NGL-05-018-104

USCEE Report 480

UNIVERSITY OF SOUTHERN CALIFORNIA

A STUDY OF SYNCHRONIZATION TECHNIQUES FOR OPTICAL COMMUNICATION SYSTEMS

R. M. Gagliardi

January 1975
Final Technical Report

Prepared for

National Aeronautics and Space Administration
Office of University Affairs
Washington, D. C. 20546

ELECTRONIC SCIENCES LABORATORY

USC Engineering

NGL-05-018-104 USCEE Report 480

A STUDY OF SYNCHRONIZATION TECHNIQUES FOR OPTICAL COMMUNICATION SYSTEMS

R. M. Gagliardi

January 1975

Final Technical Report

Prepared for

National Aeronautics and Space Administration
Office of University Affairs
Washington, D. C. 20546

This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGL-05-018-104. This grant was part of the research program at NASA's Goddard Space Flight Center, Greenbelt, Maryland.

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
2.	PROGRAM OBJECTIVE	2
3.	SUMMARY OF PROGRAM ACCOMPLISHMENTS	4
4.	PROGRAM DOCUMENT LIST	5
5.	APPENDIX (SELECTED CONTRACT REPORTS)	11

1. INTRODUCTION

This document is a final report of work done in the Department of Electrical Engineering at the University of Southern California for the National Aeronautics and Space Administration, in the area of optical communications. The work effort was carried out under the guidance of Professor Robert M. Gagliardi of the Electrical Engineering Department, and covered an extended period commencing in 1969 and ending January 31, 1975. The work was initiated as a joint research effort between the University of Southern California and NASA's Electronic Research Center in Cambridge, Massachusetts. The work was later monitored by the Electro-optics Division at the Goddard Spaceflight Center at Greenbelt, Maryland. The contract was funded by NASA's Office of University Affairs under Grant NGL 05-018-104.

The objective of the program was to study synchronization techniques and related topics in the design of high data rate, deep space, optical communication systems. The research was solely analytical in nature and was divided into two basic categories. The first involves tasks with direct application to the time synchronization problem, while the second involves related areas also being studied under the grant. The study was to indicate design procedures, assess system performance and predict future areas of needed study in synthesizing and improving digital optical systems.

This final report reviews the program objectives, the significant results, and the published research work generated during the program tenure.

2. PROGRAM OBJECTIVE

This study program was initiated in December 1968 at a time when NASA was vitally interested in developing a high data rate, deep space, optical communication system. The primary mode of operation was to be direct detection digital transmission, with interest in possible block encoding to achieve improved data rates. Use of narrow pulsed optical sources was expected to be the principle signalling format. At that time, some questions existed concerning the ability to time synchronize low duty cycle optical systems for bit and word detection. For these reasons the study program was initiated.

The specific work tasks of the program were:

1) To determine the effect of timing errors in narrow pulsed digital optical systems. This task would allow for a determination of the required timing needed in system design in order to maintain necessary temporal coherence. At the time of program commencement accurate statistical models for optical detection were only partially known. Thus, a subtask here was the development of usable system models for analysis of bit and word error probabilities for both perfect and imperfectly timed systems.

2) To determine the accuracy to which well known microwave timing systems can be operated in a low powered optical system.

3) To derive improved tracking systems for the optical channel. Also, to determine the degree of improvement that can be expected by

these newer systems, and possible upper bounds to tracking performance. This would allow comparison to present state of the art systems currently in existence, for a cost effective study of redesigning synchronization systems.

Other areas of interest closely related to the above primary tasks were also to be considered:

4) An establishment of a usable photodetector mathematical model for application to the analysis and design of performance in a communication receiver.

5) A study of the application of multi-level block encoding to the optical transmissions of digital data, and possible improvements in transmitted information rates.

3. SUMMARY OF PROGRAM ACCOMPLISHMENTS AND RESULTS

Since the study effort was solely analytical in nature, the results of the program were technical reports, summarizing the achieved milestones. The program produced a total of 11 technical reports, 12 published papers, and 2 Ph.D. dissertations. The key accomplishments of the program are summarized below. References refer to the listed reports in Section 4, where the stated results are documented.

1) Developed an accurate mathematical model of the photodetecting receiver and its statistical properties for use in digital receiver design. Specifically, a detailed study was made of the detector shot noise process and its interrelation with the counting processes that govern it. Investigation of the conditional Poisson counting process (referred to more recently as a doubly stochastic Poisson process) was made in depth, exploring the relation of Poisson, Laguerre, and Bose-Einstein counting. The relation of optical shot noise to Gaussian processes was studied. References - Section 4.1 [4, 6, 7]; Section 4.2 [3, 4, 6].

2) Studied the pulse position modulated (PPM) mode of optical digital transmission, showing its optimality and practical system implementation. The results were extended to block encoded systems and resulting error probabilities were derived. A computer program was developed for computation of PPM system performance under all possible operating conditions. Section 4.1 [1, 2, 3]; Section 4.2 [1, 2, 6, 9].

3) Determined the effects of timing errors in both PPM and on-off keyed digital systems. Section 4.1 [8]; Section 4.2 [5, 7].

4) Determined the ability to time and phase lock in phase and pulse tracking subsystems following photodetection in direct detection optical systems. Relations between tracking errors and operating signal to noise ratio were developed. Section 4.1 [5, 9]; Section 4.2 [11]; Section 4.3 [1].

5) Determined the optimal tracking system for optical systems. The approach here was to invoke the use of estimation theory and treat the tracking problem as one of estimating arrival time of a synchronizing signal. The optimal tracker was then determined as an optimal estimator of arrival time. This also allowed for a study of signal waveshape for best obtaining the timing information. Section 4.1 [10]; Section 4.2 [10].

6) Investigated digital signalling procedures other than PPM that aid in overcoming the time accuracy problem. Although PPM systems are optimal for perfectly timed systems, it was shown that use of noncoherent, multi-level frequency shift keyed systems using harmonic square waves are more efficient at high data rates. This study effort represents a new area of research not included in the task statement. Studies in this area are not completed. Section 4.1 [11]; Section 4.2 [12]; Section 4.3 [2].

7) Extended the problem of time tracking to spatial pointing, acquisition, and spatial tracking. Developed the relationship among power levels, pointing accuracy, performance, and acquisition times in locating spatially positioned transmitters. Computer programs have

been developed for this purpose. Studies in this area are still under investigation and research has not been completed.

There were not patents or inventions produced from this research.

4. PROGRAM DOCUMENT LIST

The following lists all research and technical reports, published papers, and documents generated from, and accredited to, this study grant.

4.1 Technical Reports

A69-10556 [1] NASA Technical Note TN D-4623, "M-ary Poisson Detection and Optical Communications," S. Karp and R. Gagliardi, June 1968.

A70-21778 [2] NASA Technical Note TN D-4814, "Design of PPM Optical Communication Systems," S. Karp and R. Gagliardi, October 1968.

[3] NASA Technical Note TN C-40, "Error Probabilities for Detection of M-ary Poisson Processes in Poisson Noise," S. Karp, G. Hurwitz and R. Gagliardi, May 1968.

A70-27418 [4] USCEE Report 334, "On the Representation of Continuous Stochastic Intensities by Poisson Shot Noite," R. Gagliardi and S. Karp, March 1969.

N70-42094 [5] USCEE Report 396, "Optical Synchronization - Phase Locking With Shot Noise Processes," R. Gagliardi and M. Haney, August 1970.

A71-17085 [6] USCEE Report 397, "Communication Theory for the Free Space Optical Channel," R. Gagliardi, S. Karp and E. O'Niell, August 1970.

X71-79821 [7] USCEE Report 401, "Counting Statistics for Extended Optical Photodetectors," R. Gagliardi and V. Farrukh, January 1971.

A72-25885 [8] USCEE Report 406, "The Effect of Timing Errors in Optical Digital Systems," R. Gagliardi, August 1971.

Not in system as of 7/14/75 [9] USCEE Report 426, "Synchronization Using Pulse Edge Detection in Optical PPM Communication Systems," R. Gagliardi, September 1972.

Not in system as of 7/14/75 [10] USCEE Report 448, "MAP Synchronization in Optical Communication Systems," R. Gagliardi, N. Mohanty, April 1973.

To be published in IEEE Trans. Theory, may 1975 [11] USCEE Report 471, "Noncoherent Detection of Periodic Optical Signals," R. Gagliardi, April 1974.

4.2 Published Papers

[1] R. Gagliardi and S. Karp, "M-ary Poisson Detection and Optical Communications," IEEE Trans. on Communication Technology, Vol. CT-17, No. 2, April 1969, pp. 208. A 69-10556

[2] S. Karp and R. Gagliardi, "The Design of PPM Optical Communication Systems," IEEE Trans. on Communication Technology, Vol. CT-17, December 1969, pp. 670. A70-21778

[3] R. Gagliardi and S. Karp, "On the Representation of a Continuous Intensity by Poisson Shot Noise," IEEE Trans. on Info. Theory, Vol. IT-16, No. 2, March 1970. A70-27418

[4] S. Karp, R. Gagliardi and E. O'Neill, "Communication Theory for the Free Space Optical Channel," Proc. of the IEEE, Vol. 58, No. 10, October 1970, pp. 1611. A71-17085

[5] R. Gagliardi, "On the Timing Problem in Optical Digita Systems," Proceedings of the International Telemetry Conference, September 1971, Washington, D. C. A72-12141

[6] R. Gagliardi, "Photon Counting and Laguerre Detection," IEEE Trans. on Info. Theory, Vol. IT-18, January 1972, pp. 208. A72-18394

[7] R. Gagliardi, "The Effect of Timing Errors in Optical Digital Systems," IEEE Trans. on Communication Technology, Vol. CT-20, No. 2, April 1972. A72-25885

[8] N. C. Mohanty, "On the Identifiability of Finite Mixtures of Laguerre Distributions," IEEE Trans. on Information Theory, Vol. IT-18, No. 4, July 1972. A73-36984

[9] N. C. Mohanty, "M-ary Laguerre Detection," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES, May 1973. A73-31734

[10] N. C. Mohanty, "Estimation of Delay of M PMM Signals in Laguerre Communications," IEEE Trans. on Communication, Vol. COM-22, No. 5, May 1974.

[11] R. Gagliardi, "Synchronization Using Pulse Edge Tracking in Optical PPM Communication Systems," IEEE Trans. on Comm., Vol. COM-22, No. 10, October 1974.

[12] R. Gagliardi, "Noncoherent Detection of Periodic Optical Pulses," IEEE Trans. on Information Theory, Vol. IT-5, May 1975 (to be published).

4.3 Ph.D. Dissertations

[1] "Phase Locked Loop Tracking of Shot Noise Processes," by George Michael Haney, presented to the Graduate School of the University of Southern California, January 1971.

[2] "Non-coherent Detection of Subcarrier Frequencies in Direct Detection Optical Communication Systems," by Richard A. Maag, presented to the Graduate School of the University of Southern California, February, 1975 (to be published).

5. APPENDIX

Reprints of most of the reports in Section 4 are included. When a report appeared as both a technical document and a published paper, only the paper was included. The reports included are listed in the following order: [numbers refer to their listing in Section 4.]

4.2 [1]
[2]
[3]
[4]
[6]
[7]
[8]
[9]
[10]

4.1 [5]
[10]
[11]

April 1973 USCEE Report 448

Interim Technical Report

A45-09601

MAP SYNCHRONIZATION IN OPTICAL COMMUNICATION SYSTEMS

R. M. Gagliardi
N. Mohanty

Department of Electrical Engineering

University of Southern California

Los Angeles, California 90007

This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-104. This grant was part of the research program at NASA's Goddard Space Flight Center, Greenbelt, Maryland.

ABSTRACT

The time synchronization problem in an optical communication system is approached as a problem of estimating the arrival time (delay variable) of a known transmitted field. Maximum aposteriori (MAP) estimation procedures are used to generate optimal estimators, with emphasis placed on their interpretation as a practical system device. Estimation variances are used to aid in the design of the transmitter signals for best synchronization. Extension is made to systems that perform separate acquisition and tracking operations during synchronization. The closely allied problem of maintaining timing during pulse position modulation is also considered. The results of this report have obvious application to optical radar and ranging systems, as well as the time synchronization problem.

Introduction

An important requirement in a successful communication system is to maintain accurate timing between transmitter and receiver. This timing is generally achieved by having the transmitter continually send a known clock signal to which the receiver can synchronize. For the system to be time locked, the receiver synchronization subsystem must determine the exact time at which the clock signal arrives. This measurement of clock arrival time can be considered a measurement of transmission delay time, which can be used to continually adjust the receiver clock relative to that of the transmitter. An analytical approach to the design of synchronization subsystems is to consider this arrival (delay) time measurement as an estimation problem. In this context, optimal estimators for measuring delay can then be implemented as practical devices for achieving synchronization.

In an optical communication system the arrival time measurement is hindered by both the quantum effects of the photodetection operation and by the reception of background noise radiation in the optical antenna. In this report the design of synchronizing subsystems in optical receivers is examined from an estimation point of view. Maximum aposteriori (MAP) estimators of delay are derived for both quantum limited and background additive operation, and their interpretation as practical subsystems are explored.

Problem Formulation

Let the timing information be sent from transmitter to receiver in the form of a known optical field $f(t, \underline{r})$ where $t, \underline{r}$ are the temporal and spatial

variables. The transmitted field is detected by the receiving system shown in Figure 1. A photodetector, having spatial area $\mathcal{A}$ normal to the beam propagation, intercepts the optical field producing the detector output signal x(t). The detected field has the intensity $|f(t-\tau, \underline{r})|^2$ where τ is the time delay during transmission. If we assume the field was transmitted at t = 0, then τ is alternatively the time of arrival of the field at the receiver. The detector output x(t) is given mathematically by the shot noise process

$$x(t) = c \sum_{m=0}^{k(0,t)} h(t-t_m) \tag{1}$$

where h(t) is the detector response function, c is a proportionality constant related to electron change and detector impedance, $\{t_m\}$ are the random location terms of the emitted photo electrons, and $k(t_1, t_2)$ is the random number emitted during (t_1, t_2). The latter is called the detected count process and in the absence of background field noise, is known to have a Poisson count probability with intensity parameter

$$m(t_1, t_2) = \int_{t_1}^{t_2} n(t-\tau)dt \tag{2}$$

where

$$n(t-\tau) = \int_{\mathcal{A}} |f(t-\tau, \underline{r})|^2 d\underline{r} \tag{3}$$

α = photodetection parameter

The function n(t) is the spatially integrated field intensity and is called the count intensity function. When bandlimited Gaussian white background

noise is present, the count over (t_1, t_2) is known to have a Laguerre count probability:

$$\text{Prob}[k(t_1, t_2)=k] = \left(\frac{1}{1+N_0}\right)\left(\frac{N_0}{1+N_0}\right)^k L_k^Q\left[\frac{m(t_1, t_2)}{N_0(1+N_0)}\right] \qquad (4)$$

where Q is the number of time-space modes observed over (t_1, t_2) and $\mathcal{A}$, and N_0 is the average noise count per mode.

The detector time process x(t) in (1) is then processed in the sync subsystem, herein considered a device that produces an estimate of the arrival time τ. This estimate can then be used to clock all subsequent receiver operations requiring transmitter synchronization (e.g. bit timing, ranging, etc). In typical system operation, this timing must be continually updated and the estimation of τ must be repeated by continually retransmitting the optical field. For this reason the optical field, and therefore the intensity n(t) in (3), is considered a periodic wavefrom in t with repetition period T. A receiver observation of T sec therefore corresponds to one period of the intensity waveform. The estimation problem is therefore one of observing over (0, T) the photo detected output due to repeated optical field producing the count intensity $n(t-\tau)$, and estimating the variable τ. Although we shall concentrate on the estimation problem over a single inverval, the resulting processing may then be repeated over subsequent intervals, making use of earlier estimates. Only maximum aposteriori (MAP) estimates are considered. The procedures of MAP estimation are discussed in References [1-3] , and the specific application to optical systems is reviewed in [4]. The pertinent equations necessary for this report are summarized in the Appendix.

MAP Estimation of Delay

The MAP estimate of τ under Poisson counting follows directly from the Appendix, with τ replacing θ. Since n(t) is periodic with period T, it can be expanded into a Fourier series at harmonics of frequency 1/T, each of which integrates to zero in the third term of (A- 5). Furthermore,

$$\frac{dn(t-\tau)}{d\tau} = - \frac{dn(t)}{dt} \Bigg|_{t \to t-\tau} \tag{5}$$

The MAP estimate $\hat{\tau}$ is then that τ for which

$$\max_{\tau} \left[\int_0^T x(t) \log[n(t-\tau)]\, dt + \log p(\tau) \right] \tag{6}$$

or that satisfying

$$\frac{p'(\hat{\tau})}{p(\hat{\tau})} = \int_0^T x(t) \left[\frac{d \log n(t)}{dt} \right]_{t-\hat{\tau}} dt \tag{7}$$

when the intensities are differentiable. The optimal estimator in (6) corresponds to determining the maximum of a bank of crosscorrelations of the detector output with all possible delay shifts of ln n(t), as shown in Figure 2a. Alternatively, the integral can be interpreted as the output at time τ of a passive filter whose input is x(t) and whose impulse response is ln n(-t), as shown in Figure 2b. The filter output at every t is then weighted by log p(t), and the value of t producing the maximum is the MAP estimate of τ.

When $p(\tau)$ is Gaussian with mean m_τ and variance σ_τ^2 then (7) is convenient to use, and takes the form

$$\hat{\tau} = \sigma_\tau^2 \int_0^T x(t) \left[\frac{d \log n(t)}{dt} \right]_{t-\hat{\tau}} dt + \frac{m_\tau}{\sigma_\tau^2} \tag{8}$$

The MAP estimate $\hat{\tau}$ appears on both sides and an explicit solution is not immediately available. However, we can interpret the integral as a correlation of the detector output with a delayed version of the bracketed expression. Hence, the MAP estimate is the value of τ which forces this right hand side to equal $\hat{\tau}$. This suggests an estimator similar to that shown in Figure 2c, employing a feedback loop to generate the proper $\hat{\tau}$ to force the loop to lock in (when $\hat{\tau}$ is correct the output of the correlator is that necessary to maintain the loop). Note the loop involves crosscorrelations with the time derivative of log n(t) and the specific form of the loop signal generator depends upon the transmitted intensity. If n(t) is a pure sinusoidal intensity the feedback loop specializes to the tan-lock loop [4]. If n(t) is periodic, but non-sinusoidal, the form of the MAP estimator loop changes. For example, let

$$n(t) = \frac{\mathcal{E}}{\sqrt{2\pi}D} e^{-(t^2/2D^2)}, \quad -T/2 \le t \le T/2 \tag{9}$$

representing a Gaussian shaped intensity pulse of width D and energy $\mathcal{E}$, extended periodically in time, as in Figure 3a. We assume T is many times larger than D so that the pulse occupies a relatively small portion of the observation interval and end effects can be neglected. For this case,

$$\frac{d \ln n(t)}{dt} = \frac{d}{dt}\left(-\frac{t^2}{2D^2}\right) = -\frac{t}{D} \tag{10}$$

and (8) becomes

$$\frac{\hat{\tau} - m_\tau}{\sigma_\tau^2} = \frac{1}{D^2}\int_0^T x(t)(t-\hat{\tau})dt \tag{11}$$

Hence,

$$\hat{\tau} = \frac{\int_0^T tx(t)dt + m_\tau\left(\frac{D}{\sigma_\tau}\right)^2}{\int_0^T x(t)dt + \left(\frac{D}{\sigma_\tau}\right)^2} \tag{12}$$

The integral in the denominator is the observed total number of counts $k(0, T)$. The numerator integral is the "mean", or "center of gravity", of the observed detector process $x(t)$. The MAP estimator therefore computes the "mean" or "center of gravity" of the shot noise locations in time and uses it in (12). In the typical situation the initial delay uncertainty is many times the pulse width so that $\sigma_\tau^2 \gg D^2$, and the MAP estimate is precisely this mean location time.

It is interesting to see how the estimator changes form as the optical pulse becomes sharper in form. Consider the pulse in Figure 3b, with its log derivative shown in Figure 3c. Equation (11) becomes instead

$$\frac{\hat{\tau} - m_\tau}{\sigma_\tau^2} = \int_{\hat{\tau}}^{\hat{\tau}+\varepsilon} x(t)dt - \int_{\hat{\tau}+D+\varepsilon}^{\hat{\tau}+D+2\varepsilon} x(t)dt \tag{13}$$

The feedback estimator now corresponds to the short term integration over the front and back end of the expected optical pulse, as the pulse is swept through the observation interval. In essence, the estimate is that value of

τ that "locks up" equal ϵ sec integrations separated by D sec, as shown in Figure 4. Effectively the detector output is being "gated", and the tracking loop that implements (13) is often called an early-late gate loop. Note that as $\epsilon \to 0$ in Figure 3b the pulse rise and fall time decreases, and the estimator integrates over a smaller portion of the observed output. Hence, as the optical pulse used for delay estimation is changed from a smooth Gaussian pulse to a sharper pulse waveform, the optimal estimator form changes from a center-of-gravity estimator to the early-late gate loop.

The dependence on intensity waveform can be further pursued by investigating the Carmer-Rao bound for delay estimation given in the Appendix. For a given density $p(\tau)$, the CRB decreases as the time integral in (A- 7) increases. Using (5) this integral can be rewritten as

$$\int_0^T \frac{[dn(t)/dt]^2}{n(t)}\,dt = \int_0^T \left(\frac{dn(t)}{dt}\right)\left(\frac{d \ln n(t)}{dt}\right)dt \qquad (14)$$

where the integral is over all t in (0, T) for which $n(t) \neq 0$. By applying the Schwartz inequality to the right integral, we note that (14) is maximized if

$$\frac{dn(t)}{dt} = \frac{d \log n(t)}{dt}, \quad n(t) \neq 0 \qquad (15)$$

in which case it becomes

$$\int_0^T \frac{dn(t)}{dt}\left(\frac{d \log n(t)}{dt}\right)dt = \int_0^T \left(\frac{dn(t)}{dt}\right)^2 dt \qquad (16)$$

Thus, the integral in (14) is bounded by the energy of the time derivative of the transmitted intensity. By applying Parcival's Theorem, we can further write

$$\int_0^T \left(\frac{dn}{dt}\right)^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |F_n(\omega)|^2 d\omega \qquad (17)$$

where $F_n(\omega)$ is the Fourier transform of n(t) over one period. The integral on the right can be interpreted as the mean squared frequency of the bandwidth of the intensity. Thus, the CRB for delay estimation is minimized if a transmitter intensity n(t) is used that satisfies (15) and has the largest mean square bandwidth in (17). The equality in (16) occurs only if n(t) = log n(t) + (constant) when $n(t) \neq 0$. This can be satisfied only if n(t) is constant whenever it is non-zero. Thus, (15) and (17) together suggest that best estimation (minimal CRB) corresponds to flat intensities, with as wide a frequency bandwidth as possible. The limit of such waveforms would be an ideal, rectangular, narrow pulse in time, although theoretically (16) is not valid for such intensities (the derivative of a pulse is not squared integrable). This pulsed intensity corresponds to transmission of a narrow burst of light and, in spite of the analytical difficulties, we intuitively expect such optical fields to indeed yield best delay estimation. (We may also note that the CRB for the intensity pulse in Figure 3b is approximately $\varepsilon/2\delta \log(\delta/AD)$, which decreases directly with ε and D, forcing the intensity to approach the ideal rectangular pulse.) Even though the rectangular intensity is not differentiable, the correlator-integrator in (6) retains its meaning as a short term integration over the pulse width, starting at each value of τ. This is often called a "sliding window" integrator, and the delay point where the window maximizes (6) is the MAP estimate. Unfortunately, this theoretically requires a search over all values of τ in (0, T), although this search time can often be reduced

by carrying out separate acquisition and tracking operations as discussed in the next section.

When background noise is present, the counts are governed by the Laguerre probabilities in (4). The estimation equations (6) and (7) for Poisson counting must then be replaced by the discrete operations:

$$\max_{\tau} \left[\sum_i (\log L_{k_i}) + \log p(\tau) \right] \tag{18}$$

and

$$\frac{p'(\tau)}{p(\tau)} = \sum_i k_i C(k_i, \tau) \left[\frac{d \log m_i(\tau)}{d\tau} \right] \tag{19}$$

where $k_i = k(t_i+\Delta t, t_i)$, $m_i(\tau) = m(t_2+\Delta t, t_i)$, Δt is the counting interval (reciprocal of the detector bandwidth), and

$$C(k_i, \tau) = 1 - \frac{L_{k_i - 1}}{L_{k_i}}$$

with the Laguerre functions having argument $m_i(\tau)/N_0(1+N_0)$. The summations represent modified forms of the correlation operations, and involve the count sequence over Δt sec intervals at the photodetector output.

Acquisition and Tracking in Pulse Delay Estimation

Let us consider the delay estimation problem using ideal rectangular pulses of width D, and let us write the delay τ in the form

$$\tau = jD + \tau_0, \quad k = \text{integer}; \quad 0 \le \tau_0 \le D \tag{20}$$

We are here dividing the delay into an integer multiple of pulse widths plus an additive excess portion τ_0. We can now show that the MAP estimate of τ can be obtained as $\hat{\tau} = \hat{j}D + \hat{\tau}_0$. That is, by simultaneously determining MAP estimates of j and τ_0 and substituting into (20). This follows since the joint MAP estimate of j and τ_0 must satisfy the simultaneous equations:

$$\frac{\partial p(\hat{j}, \hat{\tau}_0/x(t))}{\partial \hat{j}} = 0$$

$$\frac{\partial p(\hat{j}, \hat{\tau}_0/x(t))}{\partial \hat{\tau}_0} = 0 \qquad (21)$$

where $p(\hat{j}, \hat{\tau}_0/x(t)) = p(\tau/x(t))$ with $\tau = \hat{j}D + \hat{\tau}_0$. On the other hand, the MAP estimate of $\tau = jD + \tau_0$ satisfies $\partial(p(\hat{\tau}/k)/\partial\hat{\tau} = 0$. However,

$$\frac{\partial p(\hat{\tau}/k)}{\partial \hat{\tau}} = \frac{\partial p(\hat{\tau}/k)}{\partial \hat{j}} \cdot \frac{d\hat{j}}{d\hat{\tau}} + \frac{\partial p(\hat{\tau}/k)}{\partial \hat{\tau}_0} \cdot \frac{d\hat{\tau}_0}{d\hat{\tau}}$$

$$= \frac{\partial p(\hat{j}D+\hat{\tau}_0/k)}{\partial \hat{j}} \cdot \frac{1}{D} + \frac{\partial p(\hat{j}D+\hat{\tau}_0/k)}{\partial \hat{\tau}_0} = 0 \qquad (22)$$

If $\hat{j}$ and $\hat{\tau}_0$ simultaneously satisfy (21), then (22) is also satisfied with $\hat{\tau} = \hat{j}D + \hat{\tau}_0$. Thus, delay estimates $\hat{\tau}$ can be obtained by estimating individually the number of pulse shifts $\hat{j}$ and the amount of excess, $\hat{\tau}_0$. The estimation of j can be considered an <u>acquisition</u> problem (acquiring which interval the pulse is in), while estimation of τ_0 can be considered a <u>tracking</u> problem (tracking the excess shifts within a pulse interval). In synchronization, the time delay τ generally does not vary more than a pulse width from one observation interval to the next. This suggests an alternative, suboptimal procedure in which we obtain first a pure MAP estimate of j alone in one interval, then using $\hat{j}$ to estimate $\hat{\tau}_0$ in the subsequent interval. The system achieves initial

acquisition first, then carries out tracking over later observation intervals. The system is easier to implement and reduces search time, but we emphasize that it generally does not yield the joint MAP estimates required in (21).

To formulate the initial acquisition problem we model the observable as a vector sequence $\underline{k}$ of counts k_i over disjoint pulse widths D in (0, T). [This is equivalent to considering the h(t) functions in (1) as rectangular of width D and sampling the shot noise x(t) every D sec.] If we assume an initial apriori joint density $p(j, \tau_0)$, then we can determine the MAP estimate of j alone from

$$\max_j p(\underline{k}, j) = \max_j \int_0^D p(\underline{k}/j, \tau_0) p(j, \tau_0) d\tau_0 \qquad (23)$$

For quantum limited operation we see that when conditioned on a particular j and τ_0, the received rectangular pulse will influence only the j and j+1 interval counts, all others producing zero counts. Thus,

$$p(\underline{k}/j, \tau_0) = \left[\frac{\left[\mathcal{E} \left(1 - \frac{\tau_0}{D} \right) \right]^{k_j}}{k_j!} e^{-\mathcal{E}\left(1 - \frac{\tau_0}{D}\right)} \right] \left[\frac{(\mathcal{E}\tau_0/D)^{k_{j+1}}}{k_{j+1}!} e^{-(\mathcal{E}\tau_0/D)} \right]$$

$$= \frac{\left[\mathcal{E} \left(1 - \frac{\tau_0}{D} \right) \right]^{k_j} \left[\frac{\mathcal{E}\tau_0}{D} \right]^{k_{j+1}}}{k_j! k_{j+1}!} e^{-\mathcal{E}} \qquad (24)$$

where $\mathcal{E}$ is the received pulse energy. The MAP estimate of j is that value at which a maximum occurs in (23). Clearly, if we observe a count sequence of which two are non-zero, (24) is maximum for the non-zero k_i for any τ_0

(i.e., $\hat{j}$ is the index of the first non-zero k_i). If only one count is non-zero it can be labelled either by $k_{\hat{j}}$ or $k_{\hat{j}+1}$, and the MAP estimate is that producing the maximum. Thus, if the q^{th} count is non-zero, we must compare:

$$p(\underline{k};\hat{j}=q) = \int_0^D \left(1 - \frac{\tau_0}{D}\right)^{k_q} p(j=q)p(\tau_0/q)d\tau_0$$

$$= p(j=q) \int_0^D \sum_{i=0}^{k_q} \binom{k_q}{i} \left(-\frac{\tau_0}{D}\right)^i p(\tau_0/q)d\tau_0$$

$$= p(j=q) \sum_{i=0}^{k_q} \binom{k_q}{i} \left(\frac{-1}{D}\right)^i m_i(q) \qquad (25)$$

to

$$p(\underline{k}, \hat{j}+1=q) = p(j=q-1) \int_0^D \left(\frac{\tau_0}{D}\right)^{k_q} p(\tau_0/q-1)d\tau$$

$$= p(j=q-1)\left(\frac{1}{D}\right)^{k_q} m_{k_q}(q-1) \qquad (26)$$

where $m_i(q)$ is the i^{th} moment of the conditional density $p(\tau_0/j=q)$. Thus, if only one count is non-zero the above moment sequences of the apriori density $p(\tau_0/j)$ must be computed to determine initial MAP acquisition. If we assume the most practical case where p(j) is uniform over the integers, and $p(\tau_0/j)$ is uniform over (0, D) [initial delay is uniformly distributed over 0, T] then $m_i(q) = D^{i+1}/i+1$ for all q, and both (25) and (26) have the value D/k+1. Thus, in the uniform case, we can equally likely select q as $\hat{j}$ or $\hat{j}+1$. If no counts are non-zero we can only estimate j from its apriori density.

Once $\hat{j}$ has been determined (initial acquisition achieved) in a particular observation interval, it can be used as the true j in subsequent observation intervals in which tracking (estimating τ_0) is accomplished. With $\hat{j}$ given,

the estimate $\hat{\tau}_0$ is that value for which d ln $p(k/\hat{j}, \hat{\tau}_0)/d\hat{\tau}_0 = 0$, or that satisfying

$$\frac{-k_{\hat{j}}}{1 - \frac{\tau_0}{D}} + k_{\hat{j}+1} \left(\frac{1}{\tau_0/D} \right) = 0 \tag{27}$$

The solution is then

$$\hat{\tau}_0 = \left(\frac{k_{\hat{j}+1}}{k_{\hat{j}+1} + k_{\hat{j}}} \right) D \tag{28}$$

Thus, estimation of delay with rectangular pulses in quantum limited detection can therefore operate by first acquiring $\hat{j}$ during one observation period, then computing (28) in the next. The latter uses the observed count ratio as the fraction of the pulse width for the excess shift. As observations are made over subsequent intervals, (28) can be continually recomputed to keep track of changes in τ_0. We emphasize that we have assumed that j does not change throughout all intervals. If for some reason the delay jumps by several pulse positions, $\hat{j}$ must be re-estimated and the delay reacquired.

The variance of the above estimator is difficult to determine explicitly since $\hat{\tau}_0$ involves a ratio of random counts. In addtion, the CRB is hampered by the non-differentiability of the pulsed intensities. However, a variance upper bound on $\hat{\tau}$ can be determined by noting that Var $\tau_0 \leq D^2$. Furthermore, if all counts are zero the variance is at most that of the apriori density on τ, σ_τ^2, if we use the mean as the delay estimate. Thus,

$$\text{Variance } \tau_0 = \sigma_\tau^2[\text{Prob } \underline{k}=0] + (\text{Var } \tau_0)[\text{Prob } \underline{k}=0]$$

$$\leq \sigma_\tau^2 e^{-\mathcal{E}} + D^2(1-e^{-\mathcal{E}}) \qquad (29)$$

This shows the estimator variance is reduced to no more than the square of the pulse width D as pulse energy $\mathcal{E} \to 0$.

When background noise is present, initial acquisition is more complicated since the non-signal intervals produce noise counts also. In this case (24) is replaced by

$$p(\underline{k}/j, \tau_0) = \frac{e^{-(\mathcal{E}/1+N_0)}}{1+N_0} \left(\frac{N_0}{1+N_0}\right)^k L_{k_j}(A) L_{k_{j+1}}(B) \qquad (30)$$

where $k = \sum k_i$, $A = \mathcal{E}(1-\tau_0/D)/N_0(1+N_0)$ and $B = \mathcal{E}\tau_0/DN_0(1+N_0)$. For a given count sequence $\underline{k}$ over a particular interval, we must determine j maximizing (23), which is equivalent to determining

$$\max_j p(j) \int_0^D L_{k_j}(A) L_{k_{j+1}}(B) p(\tau_0/j) d\tau_0 \qquad (31)$$

Unfortunately, this mazimization must be found after integration over τ_0. However, we note that in comparing two different pair of indices, say (j_1, j_2) and (j_3, j_4), maximization of (30) is equivalent to comparing

$$\frac{p(\underline{k}, j_1)}{p(\underline{k}, j_3)} = \frac{\int_0^D L_{k_{j_1}}(A) L_{k_{j_2}}(B) p(\tau_0/j_1) d\tau_0}{\int_0^D L_{k_{j_3}}(A) L_{k_{j_4}}(B) p(\tau_0/j_3) d\tau_0} \gtrless 1 \qquad (32)$$

when each j is equally likely. We now see that for any τ_0 density, if

$j_1 > j_3$ and $j_2 > j_4$, then (32) exceeds one, due to the positiveness and monotonicity of Laguerre functions with their indicies. Thus, if any pair of successive counts are each greater than the corresponding members of any other pair of counts, the optimal estimate of $\hat{j}$ is always the index of the first of the larger. If no one pair dominates any other pair in this way, then one must resort to integrating first in (31). When τ_0 does not depend on j, and is uniformly distributed over D, the integration in (31) can be performed, using the identity:

$$\int_0^y L_m^Q(y-x)L_n(x)dx = L_{m+n+1}^{Q-1}(y) \qquad (33)$$

After substituting, and using again the monotonicity of the Laguerre functions, (31) becomes

$$\max_j \left\{ L_{k_j+k_{j+1}}^{Q-1}\left[\frac{\mathcal{E}}{N_0(1+N_0)}\right]\right\} = \max_j \{k_j+k_{j+1}\} \qquad (34)$$

Thus, $\hat{j}$ is the index of the pair of consecutive counts having the largest sum, and initial acquisition is achieved by determining the maximal consecutive count pair.

Lastly, we point out that the well accepted procedure of basing initial acquisition on the largest of the counts (selecting $\hat{j}$ as that j for which k_j is maximum) is equivalent to an assumption that $\tau_0 = 0$. For then B = 0 and A does not depend on τ_0 in (31), and maximization over j is equivalent to maximization over k_j.

Delay Tracking in PPM Digital Systems

A problem closely related to pulse delay estimation in synchronization occurs when considering the tracking of pulse shifts in an optical PPM system.

In this operation an optical pulse D sec wide is sent in one of M possible D sec time intervals, and a random time-shift τ_0 is added during transmission, independent of which pulse position is used. This added shift will cause PPM detection errors if not compensated, [5]. A sync subsystem of the receiver attempts to measure the added shift during each word interval for proper receiver compensation. This measurement must be made, however, without regard to the pulse position modulation. Thus during each word interval the transmitted pulse arrives with a total delay $\tau = jD + \tau_0$ as before, where j is the integer position due to the modulation and τ_0 is the added excess delay during transmission. The tracking problem can be formulated as one of estimating τ_0 in the presence of the parameter j. Because of the position modulation, j must be considered independent from one observation interval to the next, and estimates of j in one interval cannot be used in subsequent intervals. Thus, during each observation of $\underline{k}$, τ_0 must be re-estimated in the presence of j. The resulting MAP tracking system for estimating τ_0 depends upon the manner in which the index j is modeled. If j is considered an unknown parameter (no apriori density specified), then the maximization over τ_0 must take into account all the possible values that j can take on. Thus, $\hat{\tau}_0$ is the value for which

$$\max_{\tau_0} p(\tau_0/\underline{k}) = \max_{j} \left[\max_{\tau_0} p(\tau_0/\underline{k}, j) \right]$$

$$= \max_{j, \tau_0} [p(\tau_0/\underline{k}, j)] \qquad (35)$$

This is equivalent to determining simultaneous maximizing values of τ_0 and j, and therefore correspond to simultaneous estimates of these parameters. In other words, the MAP trackers must estimate both parameters each though only the estimate of τ_0 is of interest. Furthermore, both estimates must be obtained during each observation, and cannot be subdivided into acquisition and tracking, if real time solutions are desired.

If a delay of one word interval is acceptable, a suboptimal tracking procedure would be one that first estimates j during the original observation; stores the observation (detector output) for one word length, then reuses the stored observables, along with the estimate $\hat{j}$, to determine τ_0, as shown in Figure 5. The estimate of j can be made using the techniques similar to initial acquisition in synchronization. The tracking system is therefore attempting to first detect which interval contains the pulse (i.e., decode the PPM word) then uses the decoded word to estimate τ_0. In the literature, this is referred to as decision-directed estimation [2] and the resulting sync systems are called data-aided trackers [6, 7].

If the word delay in data-aided systems in prohibited an alternative scheme is that shown in Figure 6. Here estimates of $\hat{\tau}_0$ are made consecutively with each successive pair of observed counts, and stored until the end of the observation interval. The estimate of j is then used to select the τ_0 corresponding to the most likely τ_0. This operation avoids the word interval delay, but requires a bank of estimators. Both these systems are of course suboptimal since they do not necessarily produce the simultaneous maximization required in (35).

If, instead of treating $\hat{j}$ as a unknown parameter, we model it as a random variable taking on the values $1, 2, 3, \ldots, M$ with equal probability, the MAP estimate of τ_0 can be obtained by averaging over these j values. Hence we write

$$\max_{\tau_0} p(\tau_0/\underline{k}) = \max_{\tau_0} \left[\sum_{j=1}^{M} p(\tau_0/\underline{k}, j) \right] \qquad (36)$$

Since each term $p(\tau_0/\underline{k}, j)$ is the conditional density of τ_0 when the pulse is transmitted in the j^{th} position, only the k_j and k_{j+1} counts are necessary to estimate τ_0. (All other counts are either zero in the quantum limited case, or contain only noise counts, when background is present.) Hence, $p(\tau_0/\underline{k}, j)$ can theoretically be computed immediately after k_j and k_{j+1} are observed. The summation in (36) is therefore a superposition of all such aposteriori densities, each delayed until the end of the observation interval. The estimate $\hat{\tau}_0$ is then made from this superposition. The system is shown in Figure 7. Note that the delaying of the aposteriori densities can be considered as modulation removal-eliminating the position shift due to PMM-and shifting the excess delay τ_0 to the end of the interval, where the estimate is made. Note that this latter estimate is <u>not</u> simply the average of the individual MAP estimates at each value of j. If it is known that τ_0 is confined to a narrow region about each pulse position, then (36) is approximately

$$\max_{\tau_0} \{p(\tau_0/\underline{k})\} \approx \max_{\tau_0} \{p(\tau_0/\underline{k}, j_{max})\} \qquad (37)$$

where j_{max} is the j maximizing $p(\tau_0/j, \underline{k})$ over all τ_0. The last term is identical to the simultaneous estimate of j and τ_0, and therefore corresponds to the optimal MAP tracker defined in (35).

REFERENCES

[1] Van Trees, H., Detection, Estimation, and Modulation Theory-Part 1 (book), Wiley, Inc., 1968.

[2] Hancock, J. and Wintz, P., Signal Detection Theory (book), McGraw-Hill Book Company, 1966, Chapter 5.

[3] Viterbi, A., Principles of Coherent Communication, (book), McGraw-Hill Book Co., 1966, Chapter 5.

[4] Gagliardi, R. and Mohanty, N., "Estimation Theory and Optical Communications," USCEE Report 446, April, 1973.

[5] Gagliardi, R., "The Effect of Timing Errors in Optical Digital Systems," IEEE Trans. on Comm. Tech., Vol.-COM-20, April, 1972.

[6] Lindsey, W. and Simon, M., "Data-Aided Carrier Tracking Loops," IEEE Trans. on Comm. Tech., Vol. COM-19, pp. 157, April, 1971.

[7] Gagliardi, R., "Synchronization Using Pulsed Edge Tracking in Optical Systems," USCEE Report 426, September, 1972.

APPENDIX

Let $\underline{k}$ be an observable vector containing a real random parameter θ, and let $p(\theta)$ be an apriori probability density on θ. The MAP estimate of θ, given an observable $\underline{k}$, is that $\hat{\theta}$ maximizing

$$\log p(\theta, \underline{k}) = \log p(k/\theta) + \log p(\theta) \tag{A-1}$$

where $p(\underline{k}/\theta)$ is the conditional density of $\underline{k}$ given the parameter θ. In optical systems, $\underline{k}$ represents the sequence of observed photoelectron counts $(k_1, k_2, \ldots)$, each observed over a Δt sec counting interval. ($\Delta t \tilde{=}$ 1/detector bandwidth). Under quantum limited operation, the conditional density is

$$p(\underline{k}/\theta) = \prod_i [\mathcal{S}_i(\theta)]^{k_i} \exp[-\mathcal{S}_i(\theta)]/k_i! \tag{A-2}$$

where $\mathcal{S}_i(\theta)$ is the count parameter over interval $(t_i, t_i+\Delta t)$:

$$\mathcal{S}_i(\theta) = \int_{t_i}^{t_i+\Delta t} n(t, \theta)dt \tag{A-3}$$

and $n(t, \theta)$ is the count intensity. The MAP estimate of θ in (A-1) is that achieving

$$\max_{\theta} \left\{ \sum [k_i \log \mathcal{S}_i(\theta) + \mathcal{S}_i(\theta)] + \log p(\theta) \right\} \tag{A-4a}$$

The solution $\hat{\theta}$ must also satisfy the extremal condition:

$$\sum_i \left[k_i \frac{\mathcal{S}_i'(\theta)}{\mathcal{S}_i(\theta)} - \mathcal{S}_i'(\theta) \right] + \frac{p'(\theta)}{p(\theta)} \Bigg|_{\theta=\hat{\theta}} = 0 \tag{A-4b}$$

where the primes denote derivatives with respect to θ. As $\Delta t \to 0$, the continuous versions of these equations can be obtained, since $S_i(\theta) \to n(t, \theta)dt$ and $k_i \to x(t)$, the detector shot noise process. Hence (A-4) becomes

$$\max_{\theta} \left[\int_0^T x(t) \log n(t, \theta)dt + \log p(\theta) - \int_0^T n(t, \theta)dt \right] \qquad \text{(A-5a)}$$

and

$$\int_0^T x(t) \left[\frac{n'(t, \hat{\theta})}{n(t, \hat{\theta})} \right] dt + \frac{p'(\hat{\theta})}{p(\hat{\theta})} - \int_0^T n'(t, \theta)dt = 0 \qquad \text{(A-5b)}$$

The Cramer-Rao Bound lower bounds the MAP estimate, and is given by

$$\text{CRB} = \left\{ -E_{k, \theta} \left[\frac{\partial^2 \log[p(k/\theta)p(\theta)]}{\partial^2 \theta} \right] \right\}^{-1} \qquad \text{(A-6)}$$

where E is the expectation operator of $\underline{k}$ and θ. Using (A-2) in (A-6) and averaging, yields

$$\text{CRB} = \left\{ E_{\theta} \left[\frac{\partial^2 \log p(\theta)}{\partial^2 \theta} + \int_0^T \frac{[n'(t, \theta)]^2}{n(t, \theta)} dt \right] \right\}^{-1} \qquad \text{(A-7)}$$

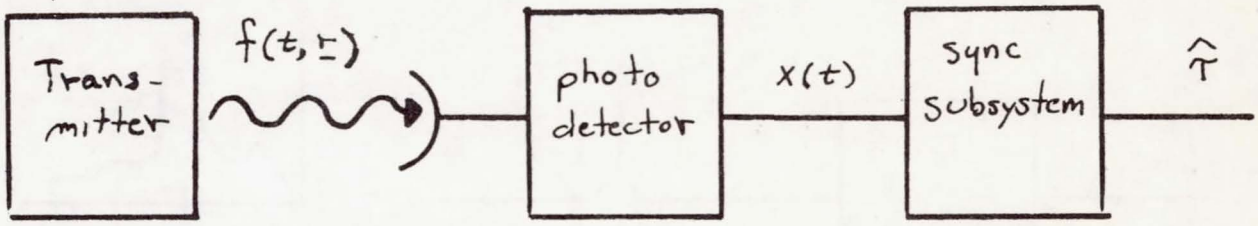
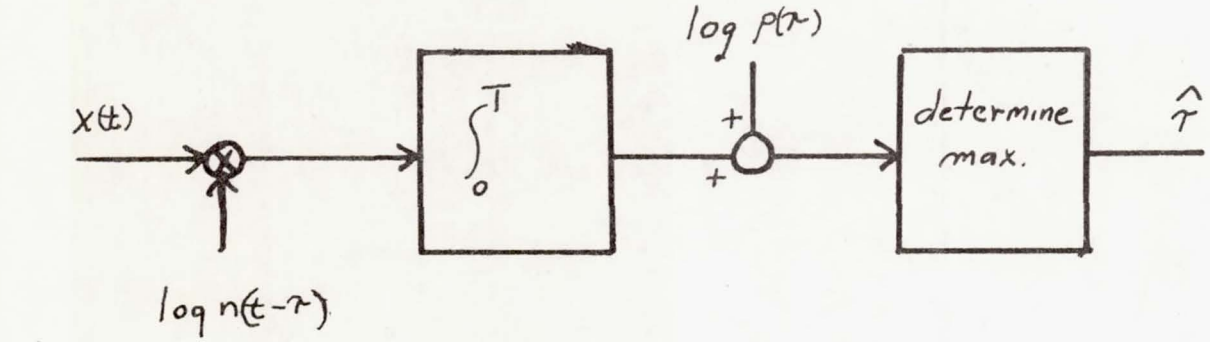
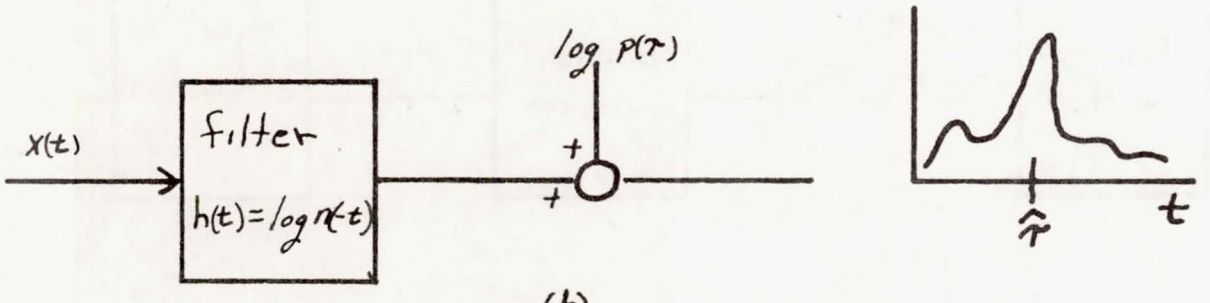


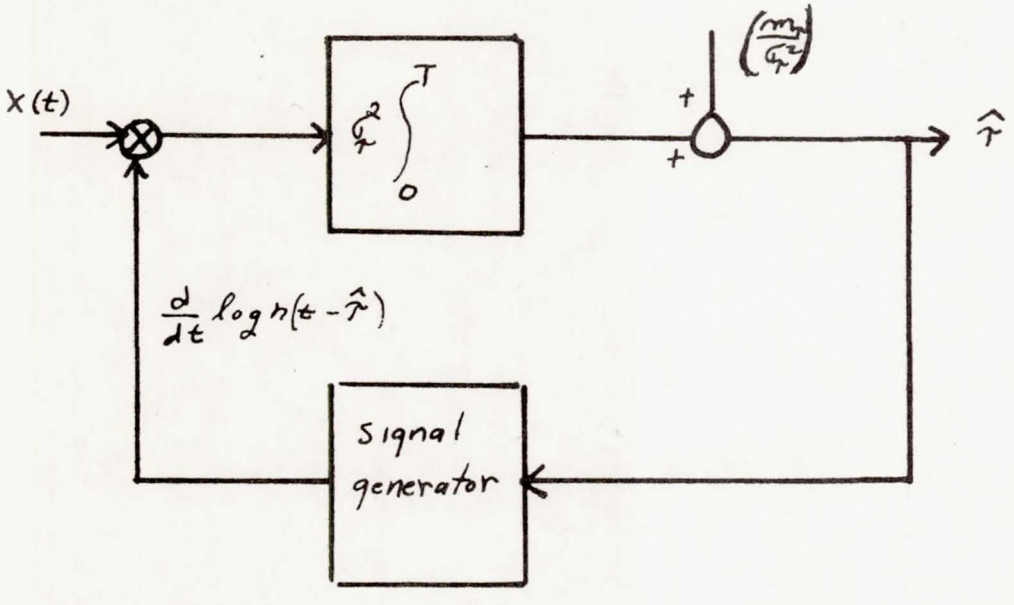
Figure 1


log p(τ)
x(t)
∫ T 0
+
+
determine max.
τ̂
log n(t-τ)

(a)


log p(τ)
x(t)
filter
h(t)=log n(-t)
+
+
τ̂
t

(b)


x(t)
∫ T 0
+
+
τ̂
d/dt log n(t-τ̂)
signal generator

(c)

Figure 2

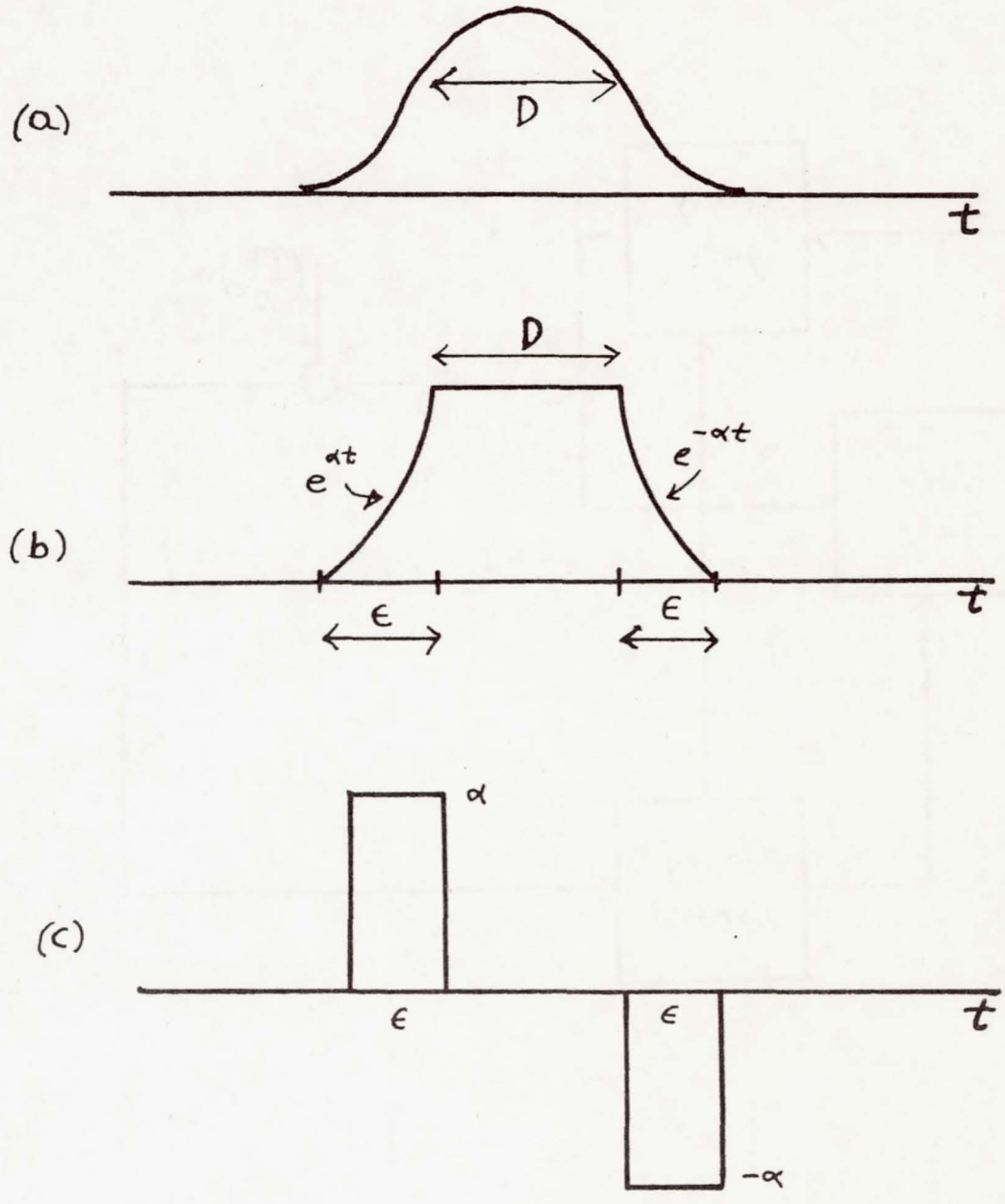

(a)
D
t
(b)
D
$e^{\alpha t}$
$e^{-\alpha t}$
t
ϵ
ϵ
(c)
α
ϵ
ϵ
t
$-\alpha$

Figure 3

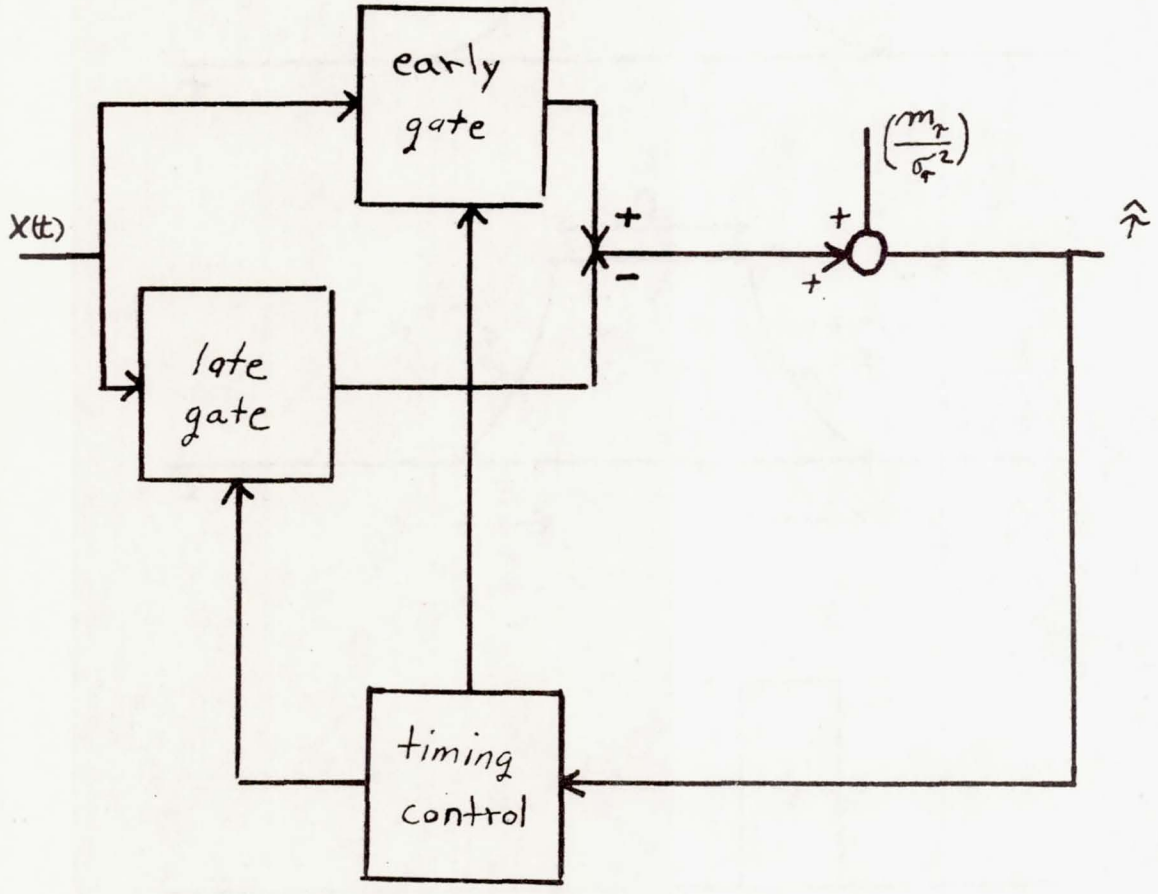

early
gate
late
gate
timing
control
X(t)
+
−
+
+
$\left(\frac{m_\tau}{\sigma_\tau^2}\right)$
$\hat{\tau}$

Figure 4

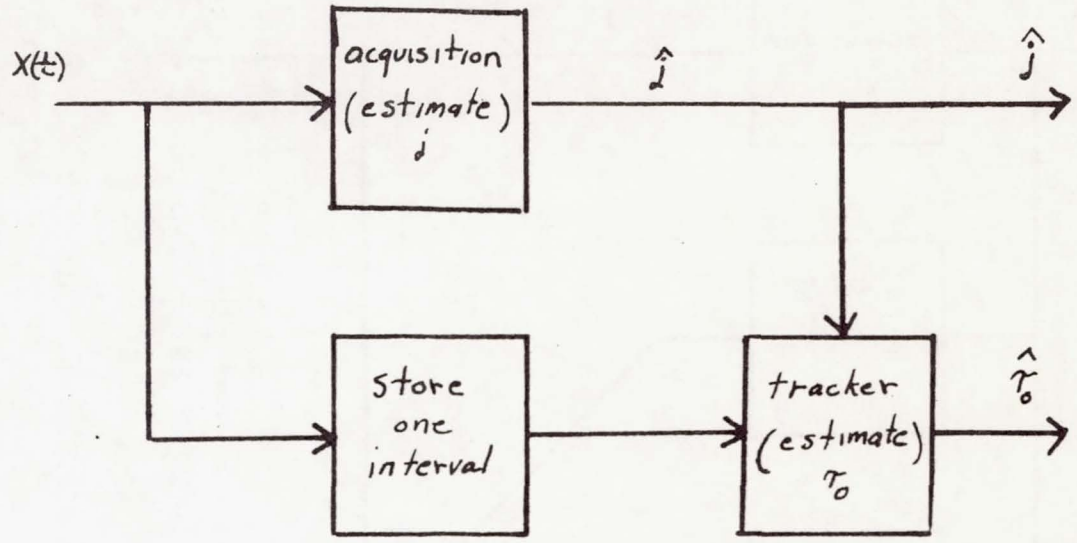

X(t)
acquisition
(estimate)
$\hat{j}$
$\hat{j}$
store
one
interval
tracker
(estimate)
τ_0
$\hat{\tau}_0$

Figure 5

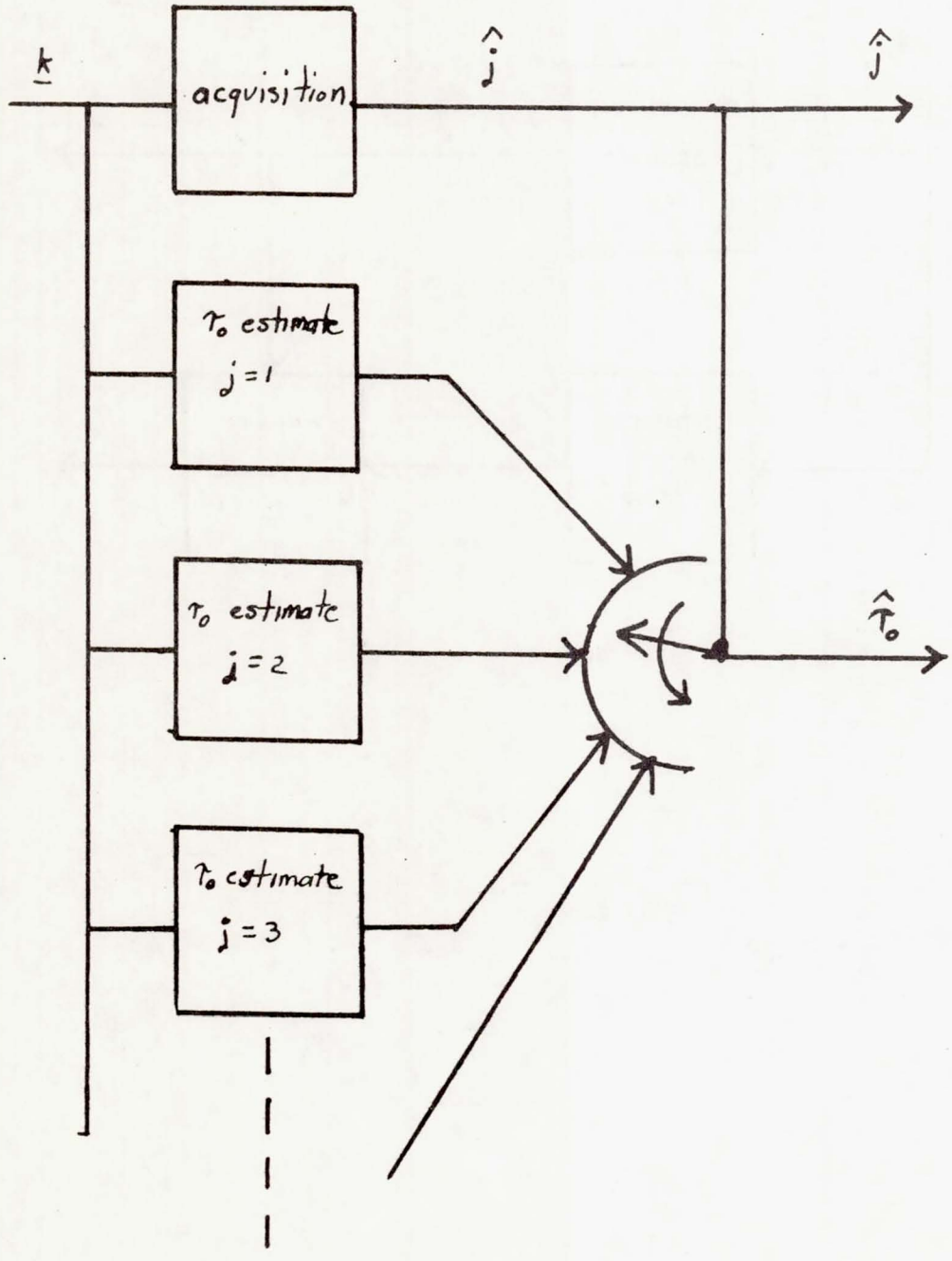

k
acquisition
$\hat{j}$
$\hat{j}$
τ_0 estimate
$j=1$
τ_0 estimate
$j=2$
τ_0 estimate
$j=3$
$\hat{\tau}_0$

Figure 6

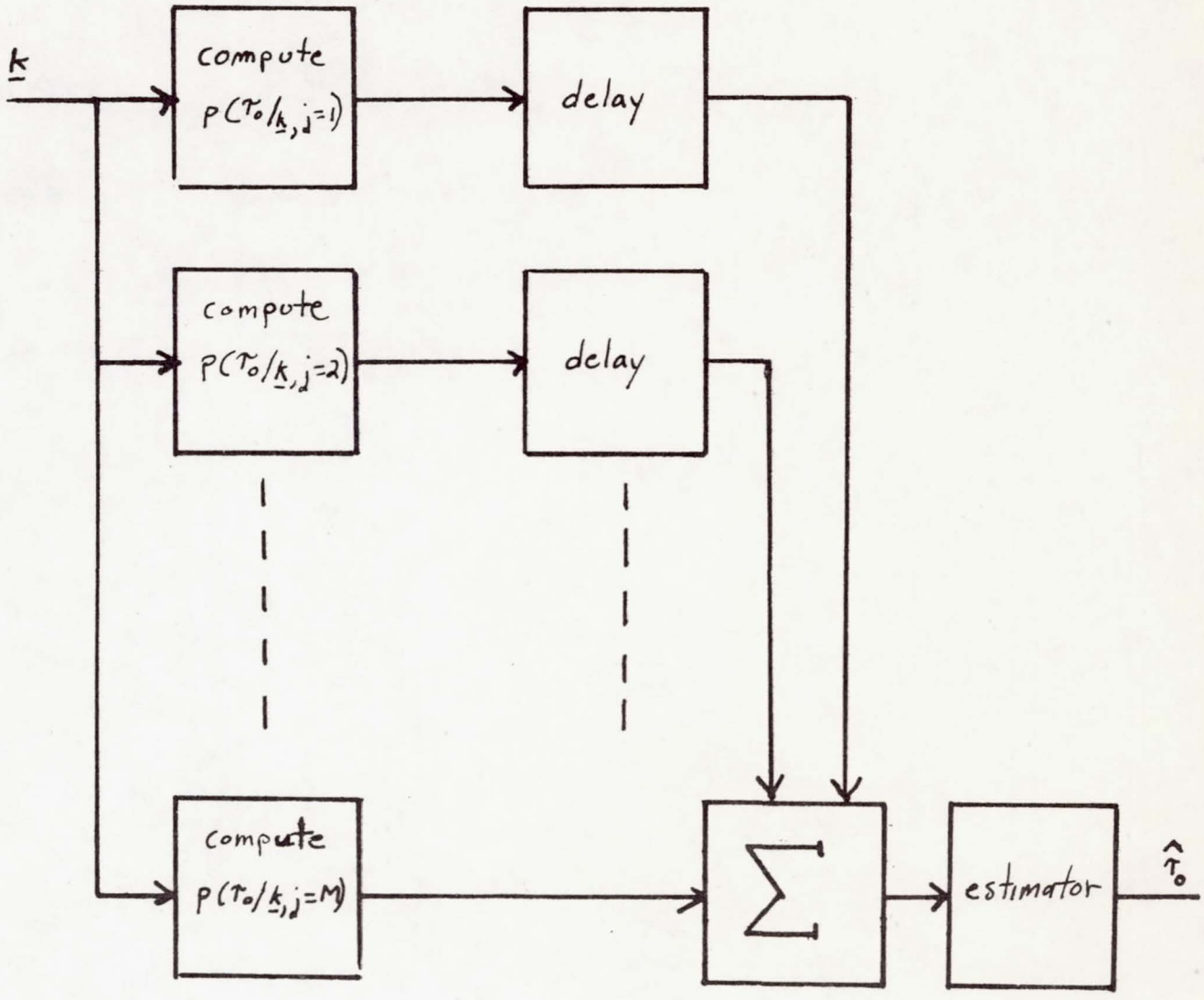

$\underline{k}$
compute
$p(\tau_0/\underline{k}, j=1)$
delay
compute
$p(\tau_0/\underline{k}, j=2)$
delay
compute
$p(\tau_0/\underline{k}, j=M)$
$\sum$
estimator
$\hat{\tau}_0$

Figure 7

April 1974

USCEE Report 471

NONCOHERENT DETECTION OF PERIODIC SIGNALS

R. Gagliardi

Interim Technical Report

Department of Electrical Engineering
University of Southern California
Los Angeles, California 90007

New Title: " Noncoherent Detection of Periodic Optical Pulses"
by R. Gagliardi
IEEE Trans on Information Theory Vol. IT-5, May 1975
To be published

This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-108-104. This grant is part of the research program at NASA's Goddard Space Flight Center, Greenbelt, Maryland.

Introduction

Various modulation techniques are presently under study for communicating digital information over an optical channel. The most common method is by the use of pulse position modulation (PPM) in which digital words are transmitted as narrow optical pulses properly located within a data frame. Such systems however are hampered by the requirement to maintain a close tolerance on timing and synchronization in order to perform detection over the narrow pulses. An alternative encoding scheme that avoids the short pulse timing problem is by the use of coded frequency division modulation (FDM). In this case information is sent as frequencies, rather than pulse positions, and the synchronization problem is relaxed. One possible implementation scheme is to transmit the digital words as bursts of square waves of different frequencies, where the length of the square wave is selected to generate sufficient energy levels for detection. The encoded square wave is used to intensity modulate the optical beam. (A square wave is used rather than a sin wave because it has maximum baseband energy in a finite time for a fixed power contraint on the optical transmitter.) Following direct (non-coherent) optical detection in the photo detector the subcarrier square wave is detected (a decision is made as to which square wave frequency is being received) in order to decode the digital word. The timing need be maintained only to within the length of the square wave signal, which is many times the length of an optical pulse in a PPM system.

It is desired to implement the optimal detector for the set of square waves. Although the bit timing problem has been considerably reduced, there still exists a time referencing problem, since the square waves will be received with random delays. Hence, coherent correlation techniques cannot be used, and the optimal noncoherent FDM square wave detector is required. Unfortunately, noncoherent detectors for waveforms that are not narrowband are not known, even for the classical additive Gaussian noise channel. In this report we present the results of an initial study to derive the optical noncoherent detector for an arbitrary periodic waveform not necessarily of the narrowband type; e.g., square waves. Attention is confined to only an additive Gaussian noise channel. The latter model is valid in an optical system when strong optical fields are detected. Future work will extend the results to the low power optical (poisson) channel.

Analysis

Classical non-coherent detection is generally understood to be the detection of a sin wave with random phase or time delay in additive gaussian noise. The problem is well documented in communication texts, and the Bayes optimal detector has been derived as both a matched envelope detector and a quadrature correlator-squaring device. These results have been expanded to include narrowband bandpass signals as well [1]. However, the extension to a general non-coherent problem involving the detection of an arbitrary periodic signal with random time delay has received little attention. Closest documentation appears in the radar literature where the problem is formulated as non-coherent detection of periodic RF pulses [2], but in all cases the narrowband assumption is imposed in order to derive an interpretable solution. Admittedly, the general non-coherent problem may not be of great practical interest because of the bandwidths required to transmit all harmonics. Also, perhaps, the complexity of the general solution may have discouraged academic pursuit. Nonetheless, in this paper the general non-coherent problem is re-examined with the objective of interpreting the processing required by the optimal detector.

Let p(t) be a general periodic, deterministic signal having period t_0 and bounded energy. The signal is observed for T seconds with a random delay τ in the presence of additive white gaussian noise r(t). The observation time T will be taken as an integer multiple of t_0 for convenience, although our results become an accurate approximation if $T >> t_0$. The observable can therefore be written

$$v(t) = p(t-\tau) + n(t) \qquad t\varepsilon(0, T) \tag{1}$$

For the non-coherent problem we assume τ is uniformly distributed over $(0, t_0)$. The optimal (Bayes) detector for the signal is desired. Mathematically, the Bayes detector is that which computes the generalized likelihood ratio Λ obtained by averaging over τ. For the observable of (1) this becomes

$$\Lambda = C\int_0^{t_0} \exp\left[\frac{2}{N_0}\int_0^T v(t)p(t-\tau)dt\right]d\tau \tag{2}$$

where N_0 is the one-sided noise level and C depends upon v(t) but not on τ. Since C can be computed without use of p(t) it is brought along simply as a constant in subsequent equations. This property of C also requires our assumption concerning the relation of observation time and signal period. Since p(t) is periodic, it admits a Fourier expansion which allows its delayed version to be written as

$$p(t-\tau) = \sum_{k=0}^{\infty} a_k \sin\left[k\left(\frac{2\pi}{t_0}\right)t + \psi_k - k\theta\right] \tag{3}$$

where (a_k, ψ_k) are the harmonic amplitudes and phases of p(t), and $\theta \triangleq 2\pi\tau/t_0$ is the uniformly distributed phase variable over $(0, 2\pi)$. The delay τ therefore introduces a random phase to each harmonic of p(t), but note that these phases are related as rational multiples of each other. Using (3) in (2), and

manipulating trigonometrically, yields

$$\Lambda = C\int_0^{2\pi} \exp\left[\sum_{k=0}^{\infty} X_k \cos k\theta + Y_k \sin k\theta\right] d\theta$$

$$= C\int_0^{2\pi} \exp\left[\sum_{k=0}^{\infty} E_k \cos(k\theta + \varphi_k)\right] d\theta \qquad (4)$$

where

$$X_k = \frac{2a_k}{N_0}\int_0^T v(t)\cos\left[k\left(\frac{2\pi}{t_0}\right)t + \psi_k\right] dt \qquad (5a)$$

$$Y_k = \frac{2a_k}{N_0}\int_0^T v(t)\sin\left[k\left(\frac{2\pi}{t_0}\right)t + \psi_k\right] dt \qquad (5b)$$

$$E_k = \left[X_k^2 + Y_k^2\right]^{\frac{1}{2}} \qquad (5c)$$

$$\varphi_k = \tan^{-1}\left[Y_k/X_k\right] \qquad (5d)$$

Here (X_k, Y_k) are the in phase and quadrature harmonic correlations, and (E_k, φ_k) are the corresponding harmonic envelope and phase variables. Unfortunately, (4) does not appear to integrate to an immediately obvious system implementation. In particular, it does not collapse down to a simple in phase and quadrature correlation with p(t) and $p[t-(t_0/2)]$, as might be conjectured from the well known bandpass case. The latter correlator would develop only if sin θ or cos θ terms factored out of every term in the exponent of (4). That this factorization does not occur in general is simply a reiteration of the fact that a single sin wave is the only periodic function satisfying the condition that shifted versions of itself are always uniquely decomposable into in phase and quadrature components.

Nevertheless, several analytical procedures are possible to reduce (4). One is to define the random variable

$$z(\theta) \triangleq \sum_{k=0}^{\infty} E_k \cos(k\theta + \varphi_k) \tag{6}$$

and to note that Λ/C is the characteristic function of z evaluated at $j\omega = 1$. Unfortunately, z is a sum of dependent random sin variables, and its probability density is not easily computed. A more fruitable procedure is to derive an infinite series solution by using the expansion

$$e^{\alpha \cos \beta} = \sum_{m=0}^{\infty} \epsilon_m I_m(\alpha) \cos m\beta \tag{7}$$

where ϵ_m is the Nueman parameter and $I_m(\alpha)$ is the m^{th} order imaginary Bessel function. When used in (4), the latter expands to

$$\Lambda = C \sum_{\underline{m}} \prod_i \epsilon_{m_i} I_{|m_i|}(E_i) \int_0^{2\pi} \cos\left[\sum_{\underline{m}} m_i i\theta + m_i \varphi_i\right] d\theta \tag{8}$$

where $\underline{m} \triangleq \{m_1, m_2, \ldots\}$ is the vector of integer coefficients m_i, $m_i \epsilon (-\infty, \infty)$. Each vector $\underline{m}$ produces a different harmonic in the integrand. However, each such harmonic will integrate to zero in (8), except for those in which

$$\sum_{i=0}^{\infty} i m_i = 0 \tag{9}$$

This reduces (8) to

$$\Lambda = C \sum_{\underline{m}(0)} \prod_{\nu} \epsilon_{m_i} I_{|m_i|}(E_i) \cos \left[\sum_{\underline{m}(0)} m_i \varphi_i \right] \tag{10}$$

where $\underline{m}(0)$ is the set of integer vectors whose components satisfy (9). The optimal detector therefore involves a search and summation over an infinite number of integer vectors. Note that the detector makes use of the envelope of each harmonic of p(t), but processes it in a rather complicated way. At this point, all that can be concluded is that the general detector involves a bank of matched envelope detectors producing $\{E_i\}$ and $\{\varphi_i\}$, followed by a complicated computer processor that instantaneously computes (10). Furthermore, the Bessel functions must be evaluated, unless one appeals to high and low signal-to-noise ratio arguments to substitute limiting forms.

Let us examine the implications of (10). Theoretically, one wonders why the optimal detector utilizes such complex processing for detection. If the harmonic random phase angles in (3) had been statistically independent of each other (i.e., $\{k\theta\}$ replaced by $\{\theta_k\}$, where the latter is an independent, uniform sequence) then the Λ obtained by averaging over the sequence of phase angles would be

$$\Lambda = C \prod_{i=1}^{\infty} I_0(E_i) \tag{11}$$

as previously reported [3]. We see that this is one term of the sum in (10). Thus the remaining terms of the sum must be taking advantage of the integer phase relation between the random phase angles. From a practical point of

view, one may also inquire if any type of physically realizable system can produce (10), precluding the use of infinitely fast computers.

A partial answer to those inquiries can be obtained by noting that (10) is reminescent of the intermodulation terms arising when a sum of carriers is passed through a nonlinearity [4]. In fact, (10) is proportional to the average, or "d.c.", value of the output of the nonlinearity e^x when impressed with the input

$$x(t) \triangleq \sum_{n=0}^{\infty} E_n \cos(nt + \varphi_n) \tag{12}$$

That is, if $y(t) \triangleq C \exp[x(t)]$, then since $x(t)$ in (12) is periodic with periodic 2π,

$$\begin{bmatrix}\text{Time average} \\ \text{of } y(t)\end{bmatrix} = \lim_{T\to\infty} \frac{C}{2T} \int_{-T}^{T} \exp[x(t)]dt = C\int_{0}^{2\pi} \exp[x(t)]dt \tag{13}$$

which is identical to the desired Λ in (4). The terms in (10) involve precisely those output harmonic terms that contribute (beat down) to this average value. The optimal processing implied is therefore used to take advantage of the phase relation among the harmonics, making use of all beat frequencies that contain useful information for detection. In the independent phase case of (11), the harmonics are not phase related and the available beat frequencies do not aid detection, on the average. Hence, only the zero order component is used. Note that the processing is not simply angle shifting each harmonic of p(t) so as to overlap in time, but instead using the nonlinearity to intentionally generate all possible beat frequencies that cause harmonic overlap.

Equation (13) also suggests a method of implementation. The receiver must generate (10), then pass it through the nonlinearity e^x, followed by averaging (low pass filtering), as shown in Figure 1. The processor generating x(t) involves determination of $\{X_n, Y_n\}$ from v(t), according to (5), then adjusting the amplitude and phase of harmonically locked oscillators, as shown in Figure 2. The computation of X_n and Y_n involve in phase and quadrature harmonic correlation over the T sec observation inverval. The overall processor would then be a bank of such harmonic subsystems, one for each signal harmonic. Since the averaging implied in (13) must be done after these correlations, Figure 1 may be interpreted as a non-real time implementation. The processor in Figure 1 can also be interpreted by comparing (12) to (6), and noting that

$$x(t) = z(\theta)\Big|_{\theta=t} \tag{14}$$

However, $z(\theta)$ is also the exponent in (2), with $\tau = t_0\theta/2\pi$. Thus

$$x(t) = \frac{2}{N_0}\int_0^T v(\rho)p\left[\rho - \left(t/\frac{2\pi}{t_0}\right)\right]d\rho \tag{15}$$

When written as above, the processor output x(t) is the output of a filter at the normalized time $t(t_0/2\pi)$, when the input is v(t) and the filter impulse response is p(-t), $(t\epsilon 0, T)$. This is simply a matched filter for the periodic signal p(t), but the filter is non-causal since p(t) is not zero for negative t. [The non-causality is indicative of the fact that all the observable over (0, T) is used to generate x(t) at any t within (0, T).] The non-causality implies

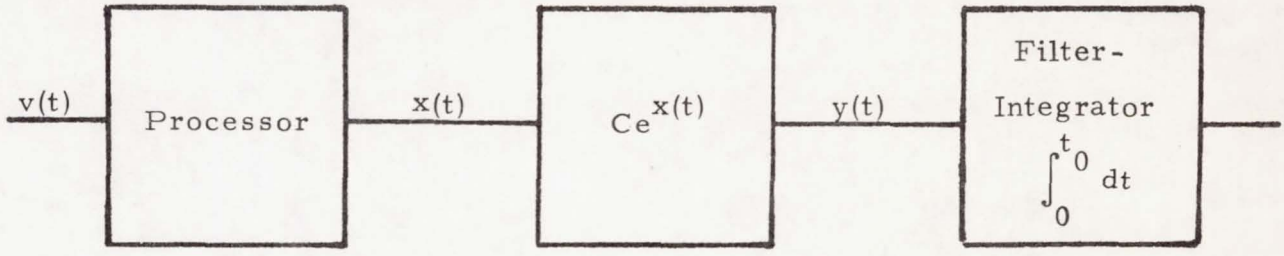


Figure 1.

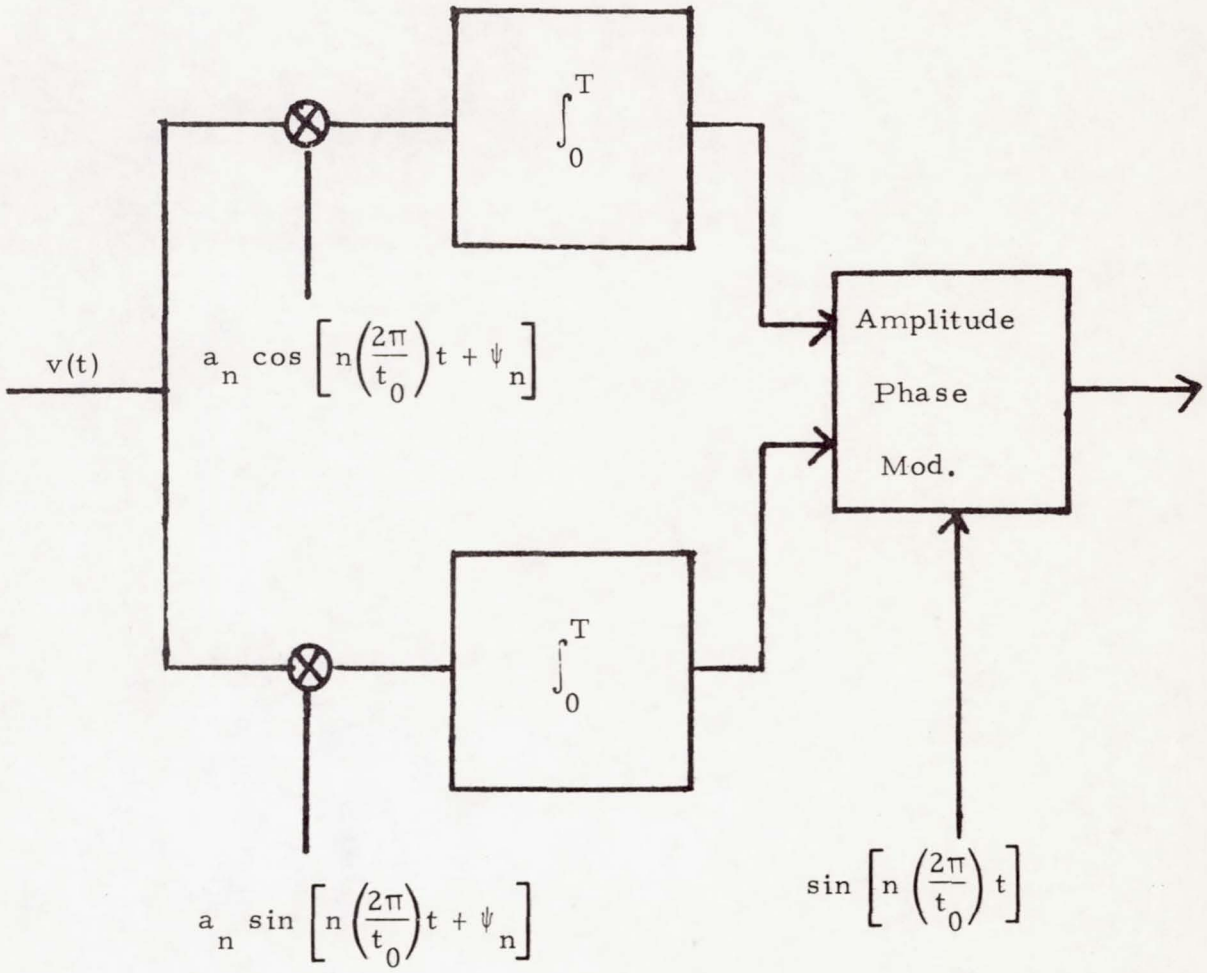

v(t)
$a_n \cos\left[n\left(\frac{2\pi}{t_0}\right)t + \psi_n\right]$
$\int_0^T$
Amplitude
Phase
Mod.
$\int_0^T$
$a_n \sin\left[n\left(\frac{2\pi}{t_0}\right)t + \psi_n\right]$
$\sin\left[n\left(\frac{2\pi}{t_0}\right)t\right]$

Figure 2.

again the non-real time implementation required for Figure 1. It is interesting that a particular non-linearity (exponential) is specified by the Bayes detector.

The extension to non-uniform densities on the delay τ can be easily accounted for in Figure 1. A non-uniform density, $\sigma(\theta)$, in the integrand of (4) would convert to a correlation rather than an integration in (13). The detector in this case would simply replace the low pass filter following the non-linearity by a correlator of y(t) and $\sigma(t)$ over the 2π sec interval. The receiver would therefore be required to locally generate this probability density as a function of t.

It may be of interest to further examine why in phase-quadrature (I-Q) correlation is not the optimal processor. The I-Q detector for an arbitrary periodic p(t) is shown in Figure 3. The input v(t) is simultaneously correlated for T sec with p(t) and $p(t-t_0/2)$, and the outputs are squared and summed. Consider the behavior of the system when only the signal portion of v(t) [i.e. $p(t-\tau)$] is impressed at the input. The output of the in phase correlator is

$$X = \int_0^T p(t-\tau)p(t)dt$$

$$= TR_{pp}(\tau) \tag{16}$$

where $R_{pp}(\tau)$ is the correlation function of p(t) evaluated at the point τ. Similarly, the quadrature correlator produces

$$Y = \int_0^T p(t-\tau)\hat{p}(t)dt$$

$$= TR_{p\hat{p}}(\tau) \tag{17}$$

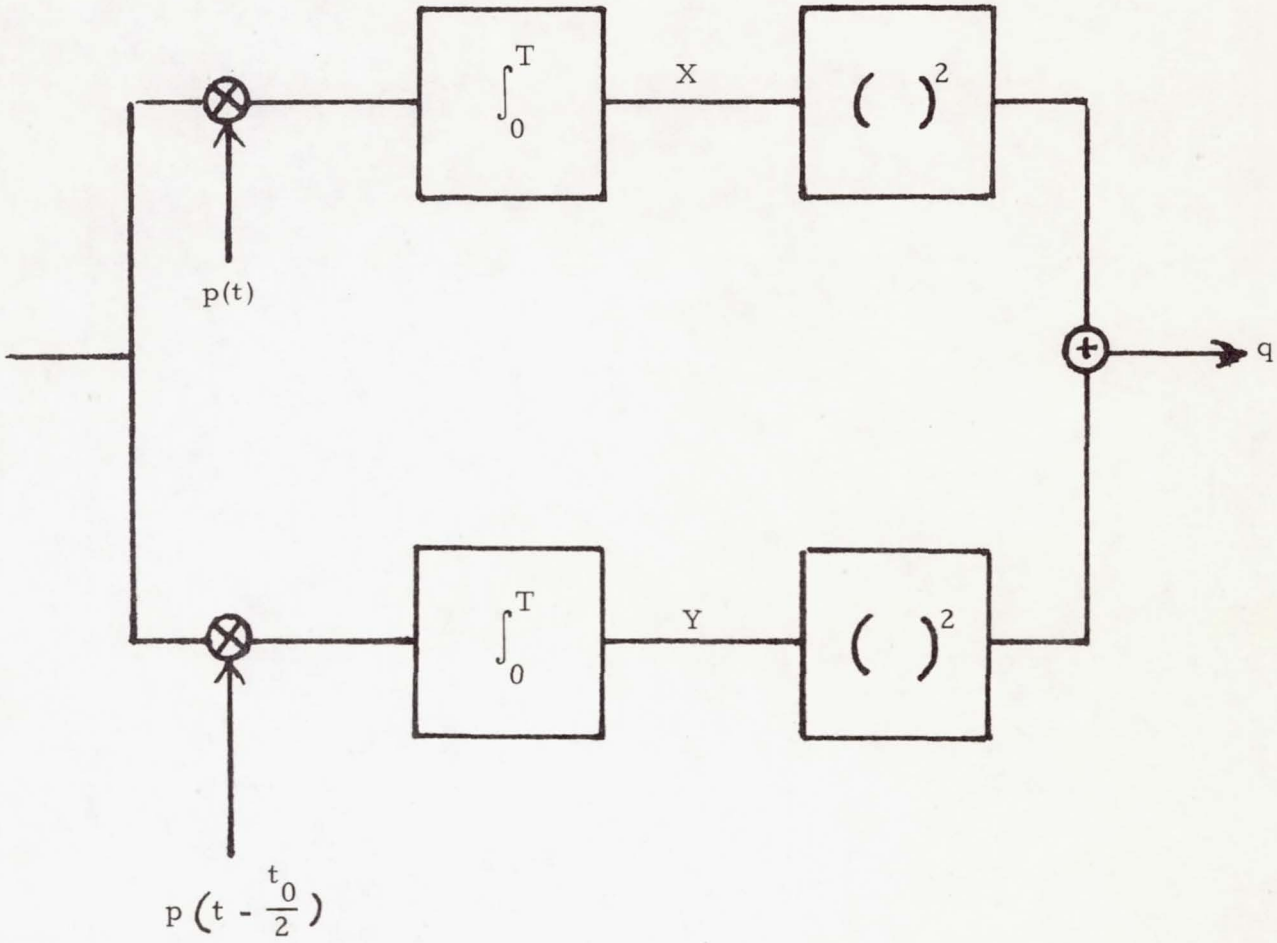


Figure 3.

where $\hat{p}(t)$ is the shifted version of p(t). Since p(t) is periodic, $\hat{p}(t)$ is also the Hilbert transform of p(t). From a well known property of such transforms [5]

$$R_{p\hat{p}}(\tau) = \hat{R}_{pp}(\tau) \tag{18}$$

Defining the complex correlation pre-envelope process $\Omega(\tau) = R_{pp}(\tau) + j\hat{R}_{pp}(\tau)$ allows us to express the I-Q correlator output as

$$\begin{aligned} q &\triangleq R^2_{pp}(\tau) + R^2_{p\hat{p}}(\tau) \\ &= |\Omega(\tau)|^2 \end{aligned} \tag{19}$$

Since $\Omega(\tau)$ is a pre-envelope process, its magnitude equals $\sqrt{2}$ times the magnitude of its real part [1, p.80]. Hence, we write q in (19) as

$$q = 2|R_{pp}(\tau)|^2 \tag{20}$$

Thus, in the noiseless case the I-Q detector always produces an output equivalent to sampling the squared correlation envelope at the delay τ. Since this τ is random it would be expected that a useful detection system should not depend on τ. The output of the I-Q detector will not depend on τ only if the envelope of the correlation function of p(t) does not depend on τ. For a pure sin wave the correlation function is a cosine wave and its envelope is indeed constant. For a narrowband bandpass p(t) the envelope is approximately constant over the range of τ [i.e., $\tau \in (0, t_0)$ and $t_0 \ll$ envelope variations]. For both of these examples the I-Q detector is in fact optimal.

However, for the general periodic function, q in (19) will depend on τ, and I-Q correlation is not a plausible detector.

REFERENCES

[1] L. E. Franks, "Signal Theory" (book) Prentice-Hall, Inc., 1969, Chapter 10.

[2] L. Wainstein, V. Zubakov, "Extraction of Signals from Noise" (book) Prentice-Hall, Inc., 1962, Chapter 6.

[3] L. Wainstein, V. Zubakov, "Extraction of Signals from Noise" (book) Prentice-Hall, Inc., 1962, p. 192.

[4] W. Davenport and W. Root, "Random Signals and Noise" (book) McGraw-Hill Book Co., 1958, p. 290.

[5] A. Papoulis, "Probability, Random Variables, and Stochastic Processes" (book) McGraw-Hill Book Co., 1965, p. 356.

USCEE Report 396

August, 1970

Optical Synchronization - Phase Locking With Shot Noise Processes

R. Gagliardi
M. Haney

Interim Technical Report

Department of Electrical Engineering
University of Southern California
Los Angeles, California 90007

This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-104. This grant is part of the research program initiated at NASA's Electronics Research Center, Cambridge, Massachusetts and continued at Goddard Space Flight Center, Greenbelt, Maryland.

Page intentionally left blank

Abstract

This report presents the results of a study effort examining time synchronization in an optical communication system. Consideration is given primarily to time locking by means of a phase lock tracking loop. Since photo-detection of an intensity modulated optical beam produces a shot noise random process at its output, synchronization analysis requires a study of phase locking with shot noise processes. A statistical analysis of tracking shot noise is presented. Of particular interest is the probability density of the tracking error, which indicates the behavior of the loop during tracking, and therefore is directly related to the ability to maintain accurate synchronization. The results of the study also have application to ranging and doppler tracking using optical systems.

Table of Contents

		Page
Abstract		i
Chapter		
1.	Introduction	1
	1.1 The Photo-Detection Model	2
	1.2 Delay Locked and Phase Locked Loops	5
2.	Error Equations For Phase Lock Loops	9
	2.1 Derivation of Loop Error Dynamics	9
	2.2 Probability Density Equations of Random Processes	14
	2.3 Probability Density Equations of Tracking Errors	18
3.	Probability Density Solutions	27
	3.1 High Electron Density Solution	27
	3.2 Higher Order Approximations	30
	3.3 Second Order Approximations	34
	3.4 Third Order Approximations	40
	3.5 Accuracy of Truncation Solutions	43
	3.6 The VCO Offset Case	46
4.	Thermal Noise and Photomultiplier Effects	50
	4.1 Additive Gaussian Thermal Noise	50
	4.2 Effect of Photomultiplication	53
5.	Second Order Loop Analyses and General Tracking Loops	57
	5.1 The Two Dimensional Smoluchowski Equation	57
	5.2 Second Order Phase Lock Loops	58
	5.3 General Delay Tracking Loops	62
	5.4 Example--Early-Late Gate Tracking	64
References		68

Chapter 1
INTRODUCTION

An important operation in communication systems is the maintenance of synchronization between transmitter and receiver. This is generally accomplished by transmitting continuously over a separate channel a known periodic waveform, and having a subsystem of the receiver continually track the waveform, thereby providing timing information for the entire receiver operation. The tracking is most typically accomplished by a delay locked loop which tracks the instantaneous time delay of the received synchronizing signal.

In an optical communication system, the synchronizing signal is often transmitted as an intensity modulated optical (laser) beam, which is photo-detected at the receiver. The subsequent timing operation is then achieved by time locking the receiver delay locked loop to the photo-detector output. Since photo-detection of an intensity modulated optical beam produces a shot noise random process at its output, the analysis of the synchronization subsystem requires careful study of the problem of time locking with shot noise input functions. In this report we present results of a study of the statistical analysis of tracking shot noise processes. Of particular interest is the probability density of the tracking error, which indicates the behavior of the loop during the tracking operation, and therefore is directly related to the ability to maintain accurate synchronization. The results of the study also have application to ranging and doppler tracking using optical systems.

1.1 The Photo-Detection Model

The overall block diagram of the sync subsystem is shown in Figure 1. The optical beam is intensity (power) modulated with a synchronizing signal. A point source photo-detection responds to the received optical radiation by producing the output shot noise process [6, 7]

$$x(t) = \sum_{m=1}^{N(0,t)} e\, h(t-t_m) \qquad (1\text{-}1)$$

where e is the electron charge, h(t) is the photo-electron waveshape in the photo-detector, t_m are the random location times of each photo-electron and N(0, t) is the number of photo-electrons occurring during the time interval (0, t). The random process N(0, t) is called the counting process of the shot noise and has a mean value given by [2, 3, 4]

$$\overline{N} = \int_0^t n(y)\, dy \qquad (1\text{-}2)$$

where

$n(t) \doteq \gamma P(t)$ = intensity of the counting process, or average rate of photo-electron occurrences.

$P(t)$ = instantaneous power in the received optical field.

γ = proportionality constant dependent upon the optical carrier frequency, Planck's constant, and the detector efficiency.

Note that the average rate of photo-electron occurrences is proportional to P(t), the power modulation on the optical beam. This means that in the case of optical synchronization, the intensity process n(t) in (1-2) is directly proportional to the synchronizing signal that power modulates the optical beam.

When the bandwidth of the photo-detector is large relative to the bandwidth of the intensity n(t), the electron functions in (1-1) can be

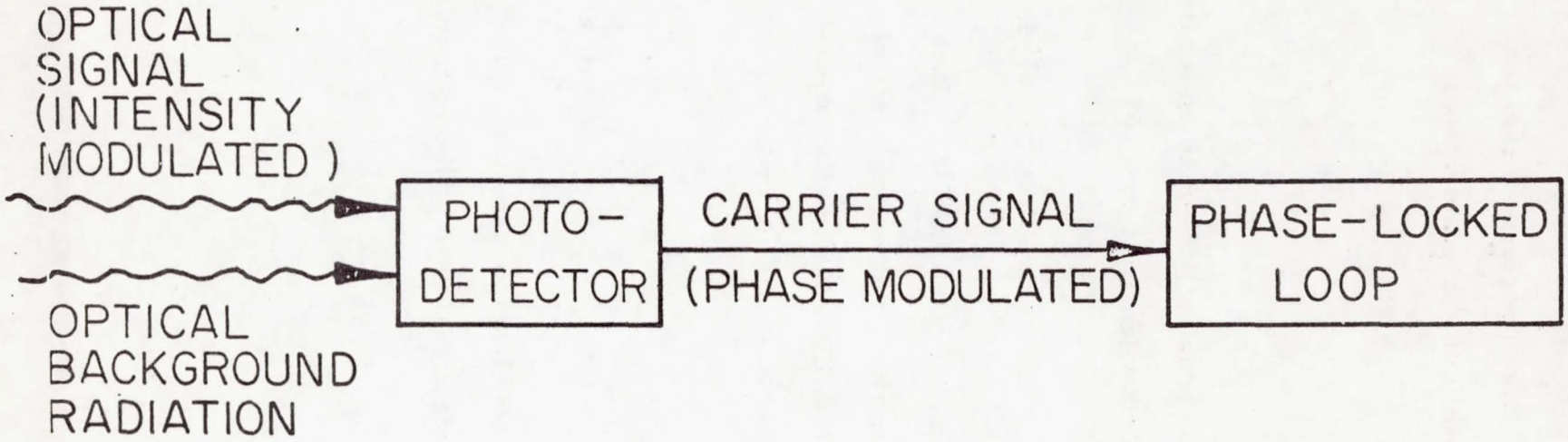

OPTICAL
SIGNAL
(INTENSITY
MODULATED)
OPTICAL
BACKGROUND
RADIATION
PHOTO-
DETECTOR
CARRIER SIGNAL
(PHASE MODULATED)
PHASE-LOCKED
LOOP

Figure 1.

considered as delta functions. In addition, $N(0,t)$ becomes a Poisson counting process [5, 6, 8], and the probability of j photo-electrons occurring in an interval $(0,t)$ is given by

$$\text{Prob}[N(0,t) = j] = \frac{\overline{N}^{j}}{j!} e^{-\overline{N}} \tag{1-3}$$

For shot noise process governed by Poisson counting, the random location times are independent, and have the probability density [5, 9, 10]

$$P(t_m) = \frac{n(t_m)}{\overline{N}} \tag{1-4}$$

where $\overline{N}$ is the average of $N(0,t)$ in (1-3) and is given in (1-2). Thus, the intensity process $n(t)$, in addition to specifying the average rate of electron occurrences, also defines the probability density of location times of the electrons. Using (1-3) and (1-4) it can be shown [5, 6] that the mean of the shot noise $x(t)$ in (1-1) is

$$[\text{mean } x(t)] = \int_0^t \delta(t-y)\, n(y)\, dy = n(t) \tag{1-5}$$

for wideband detectors. Hence, the mean of the photo-detector output in Figure 1 corresponds to the synchronizing signal used at the transmitter.

1.2 Delay Locked and Phase Locked Loops

A delay locked loop is a feedback tracking system used to time lock a locally generated periodic signal to the received periodic synchronizing signal. During each period, the two signals are time compared, and differences in timing generate error voltages that are fed back to control the timing of the local signal generator. The choice of signals at the transmitter and receiver determine the sensitivity of the error voltage to the timing difference. When the two signals are exactly in step during each period, the error voltage is zero, and the local signal remains time synchronized with the received sync signal. When this occurs, the local signal generator is producing a clean, time locked signal that can be used for timing in the remainder of the receiver. Instantaneous error voltages due to input noise represent random timing errors between the two signals, and therefore appear as synchronization errors in the receiver operation.

When the synchronizing and local signal are taken as sinusoids, the delay locked loop is called a phase lock loop [1] (since timing errors can be directly related to phase errors in the sinusoids). In phase lock loops, the signal generator is simply a voltage controlled oscillator (VCO), and the timing difference is produced in a filtered frequency mixer, as shown in Figure 2. The phase variation on the synchronizing sinusoid is then the phase signal that is to be tracked by the loop. For example, if the synchronizing signal were taken as $\sin[\omega_j t + \theta_1(t)]$, then the loop must generate an error voltage that drives the local VCO in accordance with $\theta_1(t)$.

The loop filter in Figure 2 smooths the error voltage for control of the VCO. The complexity of the loop, and of the associated analysis,

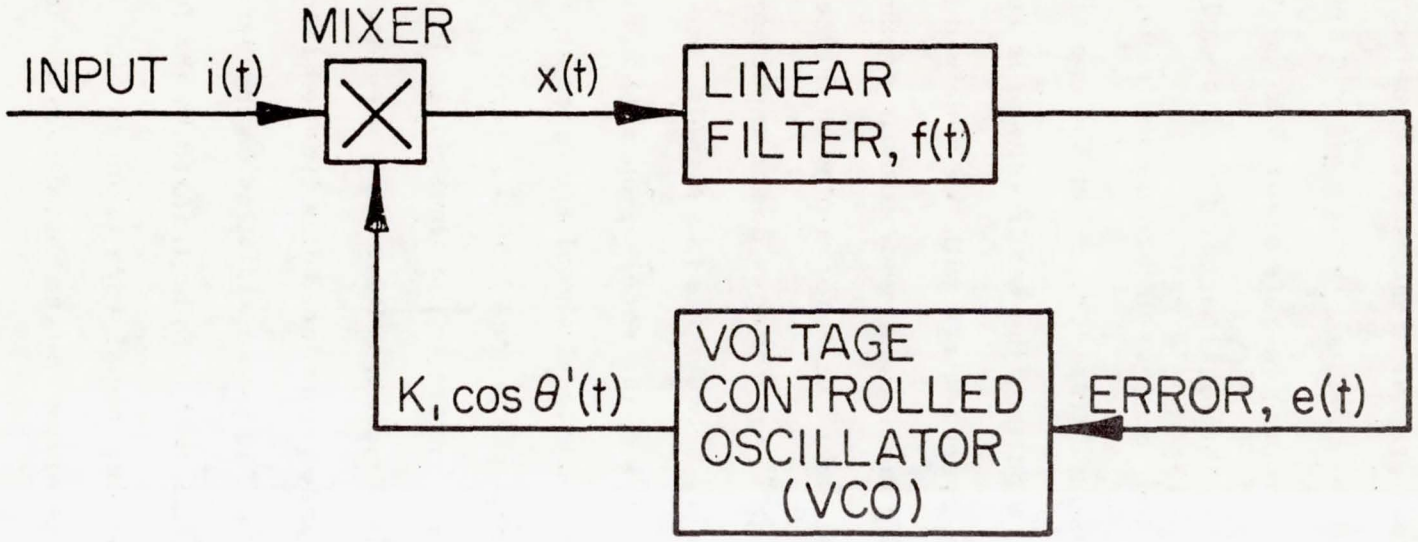

MIXER
INPUT i(t)
x(t)
LINEAR
FILTER, f(t)
VOLTAGE
CONTROLLED
OSCILLATOR
(VCO)
K, cos θ'(t)
ERROR, e(t)

Figure 2.

is determined by the type of filtering used. For a first order loop, the filter is removed and the mixer error signal feeds directly the VCO. A second order loop is produced if the loop filter effectively produces an integration. Higher order loops are generated by introducing more filter integration.

The loop mixer simply "beats" together the input and VCO sinusoid. Since the mixer is inherently bandlimited, only baseband frequencies are produced at the mixer output, while harmonics of the VCO center frequency are eliminated.

The error voltage in a phase lock loop is directly related to the phase difference between the VCO and the loop input signal at each instant of time. Hence, analytical measures of loop performance can be obtained through derivation of the loop error equations. Though these equations are generally nonlinear, the response of the loop to a "clean" synchronizing signal can usually be determined using basic nonlinear feedback analysis. Typically, the loop "pulls into" lock and the steady state loop error is driven to zero, or else the system is unstable and the loop "falls out" of lock. On the other hand, when the loop input is stochastic, the loop error responds in a random manner. In this case one can only describe the error statistically by its probability density. The derivation of this density, which is generally non-stationary, is complicated by the non-linearity of the loop. Often, we resort to a steady state density as an indication of the statistical loop behavior. The steady state variance of the loop error is then a direct indication of the phase error caused by the randomness of the input.

In the past [1, 11] the above analytical procedures have been extensively applied to the case where the input randomness is due to

additive Gaussian noise; i.e., the loop input is composed of the sum of a clean synchronizing signal plus additive Gaussian noise. However, in the optical model of Figure 1, the randomness at the loop input is due to the shot noise nature of the photo-detector output. The remainder of this report is devoted to an investigation of the loop phase error when the phase lock loop in Figure 2 is forced by the input shot noise process in (1-1).

Chapter 2

ERROR EQUATIONS FOR PHASE LOCK LOOPS

In this chapter we analytically investigate the ability of a phase lock loop to lock to a synchronizing signal that has been optically transmitted and photo-detected. Mathematically, the basic problem is that of determining the behavior of a phase lock loop when its input is a shot noise process having the synchronizing signal as its intensity process. In the following section, we derive the dynamical equations that describe the evolution of the phase error for such a system.

2.1 Derivation of Loop Error Dynamics

Consider the system shown in Figure 2 where the loop input function is the shot noise process at the wideband photo-detector output, given by (1-1):

$$x(t) = \sum_{m=1}^{N(0,t)} e\,\delta(t-t_m) \quad . \tag{2-1}$$

Here, e is the electron charge, $\{t_m\}$ are the random location times, $\delta(t)$ the electron functions and $N(0,t)$ is the shot noise counting process having intensity

$$n_s(t) = A\{1 + b\sin[\omega_s t + \theta_1(t)]\} \tag{2-2}$$

The above is proportional to the transmitted intensity modulation and represents the synchronizing signal. In (2-2), ω_s is the synchronizing frequency, b is the modulation index, $\theta_1(t)$ is the phase (time delay) variation on the synchronizing signal that is to be instantaneously tracked by the loop and A is the average value of $n_s(t)$. Recall from (1-2) that $n_s(t)$ can equivalently be interpreted as the rate of electron occurrences

in the photo-detector, so that A represents the average number of electrons produced per sec.

The VCO output in Figure 2 is represented by

$$\text{VCO output} = k_1 \cos[\omega_0 t + \theta_2(t)] \tag{2-3}$$

where k_1 is the VCO gain, ω_0 is the VCO rest frequency, and $\theta_2(t)$ its phase variation. The loop phase error is defined as the phase difference between the synchronizing signal phase and the loop VCO phase, and therefore is

$$\Phi(t) \equiv [\omega_s t + \theta_1(t)] - [\omega_0 t + \theta_2(t)] = (\omega_s - \omega_0)t + \theta_1(t) - \theta_2(t) \ . \tag{2-4}$$

The loop mixer output is then

$$e_m(t) = x(t)\,[\text{VCO output}]$$

$$= k_1 \cos[\omega_0 t + \theta_2(t)] \left[\sum_{m=1}^{N(0,t)} e\,\delta(t - t_m) \right] \tag{2-5}$$

and the loop filter output is

$$e_f(t) = \int_0^t e_m(\tau)\, f(t-\tau)\, d\tau \ . \tag{2-6}$$

The VCO output phase responds to the VCO input control voltage $e_f(t)$ through the linear relation

$$\frac{d\theta_2(t)}{dt} = k_2\, e_f(t) \tag{2-7}$$

with k_2 a constant of proportionality. From (2-4) we have, upon differentiating,

$$\frac{d\Phi(t)}{dt} = (\omega_s - \omega_0) + \frac{d\theta_1}{dt} - \frac{d\theta_2}{dt}$$

$$= (\omega_s - \omega_0) + \frac{d\theta_1}{dt} - k_2 e_f(t) . \qquad (2-8)$$

The term $(\omega - \omega_0)$ is the difference between the input synchronizing frequency and the VCO rest frequency and is called the frequency "offset" of the loop. Substitution from (2-6) then yields

$$\frac{d\Phi}{dt} = (\omega_s - \omega_0) + \frac{d\theta_1}{dt} - e\,k \int_0^t f(t-\tau) \cos[\omega_s \tau + \theta_1(\tau) - \Phi(\tau)] \cdot \sum_{m=1}^{N(0,\tau)} \delta(\tau - t_m)\, d\tau \qquad (2-9)$$

where $k = k_1 k_2$, and can be interpreted as the total gain around the loop. Equation (2-9) is then the stochastic integro-differential equation that describes the behavior of the loop phase error in terms of the input signal and loop parameters. Note that it is a non-linear equation with $\Phi(t)$ appearing on both sides of the equation. The input shot noise and the phase variation of the transmitted synchronizing signal play the role of "forcing" functions in the generation of the error process. Since the input shot noise contains random parameters, the solution for $\Phi(t)$ necessarily evolves as a stochastic process.

We ultimately will be interested in the statistical properties of the phase error. We may however note that a sample expression for the mean of $\Phi(t)$ in (2-9) can be generated, which may be useful in signal design. If we average both sides of (2-9) and interchange averaging and differentiation on the left, we see that

$$\frac{\partial \bar{\Phi}(t)}{\partial t} = \left[(\omega_s - \omega_0) + \frac{d\theta_1}{dt}\right] - e\,k \int_{-\infty}^{t} f(t-\tau) \cdot \cdot \left\{E \cos[\omega_s \tau + \theta_1 - \Phi(\tau)] \sum_{m=1}^{N} h(\tau - t_m)\right\} d\tau \qquad (2\text{-}10)$$

where $\bar{\Phi}(t)$ is the mean of $\Phi(t)$. The averaging in the integrand can be carried out by using the conditional expectations:

$$E[\ \ ] = E_{\Phi}\left\{E_{t_m, N|\Phi}[\ \ ]\right\} . \qquad (2\text{-}11)$$

The inner expectation involves only the average of the shot noise, which is given in (1-5) as

$$E \sum_{m=1}^{N} \delta(t - t_m) = n_s(t) . \qquad (2\text{-}12)$$

Substitution into (2-11), allows us to rewrite the braces in (2-10) as

$$E_{\Phi}\left\{\sin[(\omega_0 - \omega_s)t + \Phi]\right\} + E_{\Phi}[\text{terms at } (\omega_0 + \omega_s)] . \qquad (2\text{-}13)$$

The loop filtering in (2-9) eliminates the sum frequency term. Hence, (2-10) becomes

$$\frac{\partial \bar{\Phi}(t)}{\partial t} = \left[(\omega_s - \omega_0) + \frac{d\theta_1}{dt}\right] - ek \int_{-\infty}^{t} f(t-\tau)\, E_{\Phi}\left\{\sin(\omega_0 - \omega_s)\tau + \Phi(\tau)\right\} d\tau \qquad (2\text{-}14)$$

The above is interesting in that it shows that if the loop is tracking frequency and phase fairly accurately (i.e., $\omega_s = \omega_0$ and $\sin(\varphi) \approx \varphi$), then (2-14) is approximately

$$\frac{\partial \bar{\Phi}(t)}{\partial t} \simeq \frac{d\theta_1}{dt} - ek \int_{-\infty}^{t} f(t-\tau)\, \bar{\Phi}(\tau)\, d\tau . \qquad (2\text{-}15)$$

This equation has the form of the deterministic tracking error produced in the linear feedback loop shown in Figure 3, when forced by the input $\theta_1(t)$. Note that the equivalent linear loop replaces the VCO by an integrator, the mixer by a subtractor, and retains the same loop filter. Hence, the loop error function in (2-9) has a mean value such that when the loop is tracking well [i.e., $|\Phi(t)| << 1$], the mean varies in time according to the error function of the linear system in Figure 3. The latter system can therefore be used to design loop filters and compute mean error performance.

For a complete statistical analysis, however, we must return to (2-9) for study. The complexity of the error process $\Phi(t)$ is exhibited even if we consider a simplified special case. For example, consider a first order loop in which the loop filter is removed. [This effectively replaces f(t) by a delta function in (2-6).] In this case, (2-9) becomes

$$\frac{d\Phi}{dt} = \left[(\omega_s - \omega_0) + \frac{d\theta_1}{dt}\right] - ek \cos[\omega_s t + \theta_1(t) - \Phi(t)] \cdot \sum_{m=1}^{N(0,t)} \delta(t-t_m) . \qquad (2\text{-}16)$$

Though simplified, (2-16) is still a non-linear differential equation involving the random loop error process $\Phi(t)$. By integrating both sides we note

$$\Phi(t) = [(\omega_s - \omega_0)t + \theta_1(t)] - ek \sum_{m=1}^{N(0,t)} \cos[\omega_s t_m + \theta_1(t_m) - \Phi(t_m)] . \qquad (2\text{-}17)$$

The second term represents a summation of random "jumps", the height of the jumps dependent upon $\Phi(t)$ itself. This identifies the process $\Phi(t)$ in (2-16) as a discontinuous, or "jump", process in which the

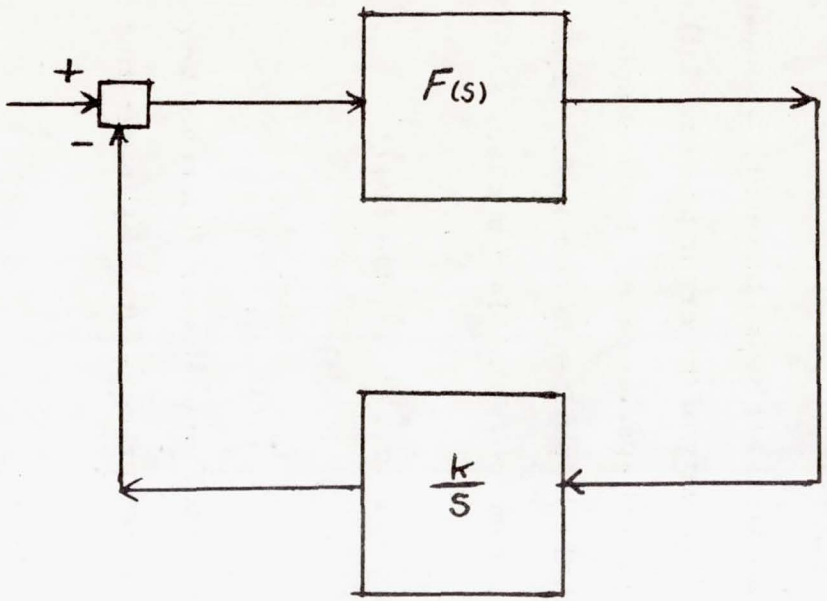


Figure 3

number of jumps are governed by the counting process $N(0,t)$. Therefore, even for this specialized case, the complexity of the error process is apparent.

In the following section we derive an equation involving the probability density of a general random process. Subsequently, we shall apply the result to the error process generated in (2-9).

2.2 Probability Density Equations of Random Processes

Let $\Phi(t)$ be a scalar random process, and let $p(\Phi_1, t_1)$ represent the probability density function (pdf) of the process at time t_1 in the variable Φ_1. Similarly, denote $p(\Phi_1, t_1 | \Phi_2, t_2)$ as the conditional pdf of $\Phi(t)$ at time t_1, given that $\Phi(t_2) = \Phi_2$ at time t_2. The pdf is then always related to the conditional pdf by

$$p(\Phi_1, t_1) = \int_{-\infty}^{\infty} p(\Phi_1, t_1 | \Phi_2, t_2)\, p(\Phi_2, t_2)\, d\Phi_2 \quad . \tag{2-18}$$

Note that the conditional density can be interpreted as a transitional density in the sense that it "converts" the pdf at time t_2 to its new density at time t_1. When $t_1 > t_2$, this transitional density essentially indicates the manner in which the pdf propagates in time.

Equation (2-18) can be rewritten in a different form for convenient interpretation and application. Define the conditional characteristic function of the random increment $\Delta\Phi = \Phi_2 - \Phi_1$ as

$$C_{\Delta}(\omega) = \int_{-\infty}^{\infty} e^{j\omega(\Phi_1 - \Phi_2)}\, p(\Phi_1, t_1 | \Phi_2, t_2)\, d\Phi_1 \quad . \tag{2-19}$$

By inverse Fourier transform

$$p(\Phi_1, t_1 | \Phi_2, t_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega(\Phi_1 - \Phi_2)}\, C_{\Delta}(\omega)\, d\omega \quad . \tag{2-20}$$

Substitution of (2-20) into (2-18) then yields

$$p(\Phi_1, t_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} p(\Phi_2, t_2)\, d\Phi_2 \int_{-\infty}^{\infty} e^{-j\omega(\Phi_1 - \Phi_2)}\, C_\Delta(\omega)\, d\omega \quad . \tag{2-21}$$

Now it is well known that the characteristic function can be expanded into moments as

$$C_\Delta(\omega) = 1 + \sum_{i=1}^{\infty} \frac{(j\omega)^i}{i!}\, m_i(\Delta\Phi) \tag{2-22}$$

where

$$m_i(\Delta\Phi) = E_{\Phi_1}[(\Phi_1 - \Phi_2)^i \mid \Phi_2] \tag{2-23}$$

is the i^{th} conditional moment of $\Delta\Phi$ given $\Phi(t_2) = \Phi_2$. [Alternatively, $m_i(\Delta\Phi)$ are the moments of the conditional pdf in (2-20).] It follows that

$$p(\Phi_1, t_1) = \sum_{i=0}^{\infty} \frac{1}{2\pi\, i!} \int_{-\infty}^{\infty} p(\Phi_2, t_2)\, d\Phi_2 \int_{-\infty}^{\infty} m_i(\Delta\Phi) e^{-j\omega(\Phi_1 - \Phi_2)}\, (j\omega)^i\, d\omega \quad . \tag{2-24}$$

But

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega(\Phi_1 - \Phi_2)}\, (j\omega)^i\, d\omega = \left(-\frac{\partial}{\partial\Phi_1}\right)^i \int_{-\infty}^{\infty} e^{-j\omega(\Phi_1 - \Phi_2)}\, d\omega$$

$$= \left(\frac{\partial}{\partial\Phi_1}\right)^i \delta(\Phi_1 - \Phi_2) \tag{2-25}$$

and (2-24) becomes

$$p(\Phi_1, t_1) = \sum_{i=0}^{\infty} \frac{1}{i!} \int_{-\infty}^{\infty} m_i(\Delta\Phi) \left(-\frac{\partial}{\partial\Phi_1}\right)^i p(\Phi_2, t_2)\, \delta(\Phi_1 - \Phi_2)\, d\Phi_2$$

$$= p(\Phi_1, t_2) + \sum_{i=1}^{\infty} \frac{1}{i!} \left(-\frac{\partial}{\partial\Phi_1}\right)^i m_i(\Delta\Phi)\, p(\Phi_1, t_2). \qquad (2\text{-}26)$$

The first term is the pdf at time $t = t_2$, and the summation represents the increment in this latter pdf to produce the pdf at $t = t_1$. If we set $t_2 = t$ and $t_1 = t + \Delta t$, then (2-26) becomes

$$p(\Phi_1, t + \Delta t) - p(\Phi_1, t) = \sum_{i=1}^{\infty} \frac{1}{i!} \left(-\frac{\partial}{\partial\Phi_1}\right)^i m_i(\Delta\Phi)\, p(\Phi_1, t_2). \qquad (2\text{-}27)$$

Dividing by Δt and passing to the limit as $\Delta t \to 0$ we obtain

$$\frac{\partial\, p(\Phi, t)}{\partial t} = \sum_{i=1}^{\infty} \frac{1}{i!} \left(-\frac{\partial}{\partial\Phi}\right)^i [K_i(\Phi)\, p(\Phi, t)], \qquad (2\text{-}28)$$

where

$$K_i(\Phi) = \lim_{\Delta t \to 0} \left[\frac{E[(\Delta\Phi)^i \mid \Phi)]}{\Delta t}\right]. \qquad (2\text{-}29)$$

Equation (2-28) is called the stochastic kinetic equation [17], or the Smoluchowski-Komogorov equation [16]. When the coefficients $K_i(\Phi)$ exist, this equation provides a relation that must be satisfied by the pdf of the process $\Phi(t)$. Note that the equation is a partial differential equation with variable coefficients, and involve all orders of derivatives. The remarkable point is that no continuity conditions on $\Phi(t)$ were required, so that the equation is valid whether $\Phi(t)$ is continuous or not. In essence, the integral equation in (2-18) has been replaced by the differential equation in (2-28). Furthermore, while one needs the complete conditional pdf to carry out (2-18), only the moments of this density are needed to derive (2-28).

The principle usefulness of the Smoluchowski equation occurs when only the first few coefficients $K_i(\Phi)$ are non-zero. In particular, if $K_i = 0$, $i \geq 3$, the resulting equation is called the Fokker-Planck equation, and has been extensively studied [1]. The Fokker-Planck equation will arise whenever the random process $\Phi(t)$ is continuous, while discontinuous processes generate all the coefficients in (2-28) [17]. We would expect this latter condition to be true for our process $\Phi(t)$, based upon our earlier discussion of the apparent jump nature of the error function. Equation (2-28) is a partial differential equation of the type

$$\frac{\partial p(\Phi, t)}{\partial t} = L_{\Phi}[p(\Phi, t)] \tag{2-30}$$

where L_{Φ} is a differential operator in Φ. The usual method for solving this type of equation is by separation of variables. In this method it is assumed that

$$p(\Phi, t) = K(t)\, p(\Phi) \tag{2-31}$$

and a solution is desired that satisfies Equation (2-30) with the appropriate initial conditions. Substitution into Equation (2-30) yields

$$\frac{1}{K(t)} \frac{dK(t)}{dt} = \frac{1}{p(\Phi)} L_{\Phi}[p(\Phi)] . \tag{2-32}$$

Since the left side depends only on t, and the right side only on Φ, they can be equal only if they equal a constant. Thus

$$\begin{aligned} &\frac{dK(t)}{dt} = cK(t) \\ &L_{\Phi}[p(\Phi)] = cp(\Phi) \end{aligned} \tag{2-33}$$

for some c if a solution is to be found by this method. Furthermore, if $\{ c_i \}$ is a set of values of c which satisfy the above, then $p(\Phi, t)$ must be of the form

$$p(\Phi, t) = \sum_i B_i(\Phi) e^{-c_i(t)} \tag{2-34}$$

where the $\{B_i(\Phi)\}$ are determined by appropriate initial conditions. Since each term of the sum approaches zero as t goes to infinity for all c_i greater than zero, the steady state solution, $p(\Phi)$ (defined as the limiting form of $p(\Phi, t)$ as $t \to \infty$), must be due to the value of $c_i = 0$. Therefore, from (2-33), the steady state solution satisfies

$$L_\Phi [\, p(\Phi)] = 0 \,. \tag{2-35}$$

Thus, the steady state solution to (2-35) (if one exists) is the solution to a differential equation obtained by setting the right hand side of (2-30) equal to zero and replacing $p(\Phi, t)$ by $p(\Phi)$.

2.3 Probability Density Equations of Loop Tracking Errors

It has been shown that a general random process has a probability density which satisfies the Kolmogorov partial differential equation. We have seen that this equation may, however, involve an infinite number of derivative terms. In this section we would like to derive the corresponding pdf equation for the phase error process of a tracking loop, governed by the dynamical equation in (2-9). To accomplish this, we must calculate the sequence of moment coefficients $K_i(\Phi)$ given by (2-29). This in turn requires determinations of the phase increment $\Delta\Phi$ of $\Phi(t)$ during the interval $(t, t + \Delta t)$.

Consider a first order phase lock loop tracking a synchronizing signal with a constant delay, following wideband photo-detection. The phase error $\Phi(t)$ then satisfies the differential equation (2-16), and has the form:

$$\frac{d\Phi}{dt} = ek \cos[\omega_0 t + \theta_1 - \Phi(t)] \cdot \sum_{m=1}^{N(0,t)} \delta(t-t_m) \tag{2-36}$$

where θ_1 is the constant phase delay. Note that the forcing function in (2-15) is zero, so that the steady state mean error is zero. The phase variation $\Delta\Phi$ is obtained by integrating $d\Phi$ from t to $t + \Delta t$. Thus, from (2-16)

$$\begin{aligned}\Delta\Phi &\equiv \int_t^{t+\Delta t} d\Phi = \int_t^{t+\Delta t} \left(\frac{d\Phi}{dt}\right) dt \\ &= -ek \int_t^{t+\Delta t} \cos[\omega_s t + \theta_1 - \Phi(t)] \cdot \sum_{m=1}^{N(0,t)} \delta(t-t_m)\, dt \\ &= -ek \sum_{m=0}^{N(\Delta t)} \cos[\omega_s t_m + \theta_1 - \Phi(t_m)]\end{aligned} \tag{2-37}$$

where $N(\Delta t)$ is the number of electron occurrences in the interval $(t, t + \Delta t)$. The above expresses the increment of the phase variation during $(t, t + \Delta t)$. Note that this variation is also a "jump process", having randomly occurring "jumps" of random heights, and that the argument of the cosine function depends upon the process $\Phi(t)$ itself (which emphasizes the non-linearity of the loop dynamics).

Now, from Equation (2-29)

$$K_n(\Phi) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E_{N,\, t_m} [(\Delta\Phi)^n | \Phi]$$

$$= \lim_{\Delta t \to 0} \frac{(-ek)^n}{\Delta t} E_{N,\, t_m/\Phi} \left[\sum_{m=1}^{N(\Delta t)} \cos^{\theta'}(t_m) \right]^n \qquad (2\text{-}38)$$

where $\theta' = [\omega_0 t + \theta_1 - \Phi]$ and the expectation is conditioned on Φ. The quantity in brackets becomes

$$\sum_{m_1=1}^{N(\Delta t)} \sum_{m_2=1}^{N(\Delta t)} \cdots \sum_{m_n=1}^{N(\Delta t)} \cos\theta'(t_{m_1}) \cos\theta'(t_{m_2}) \cdots \cos\theta'(t_{m_n})$$

which is

$$\sum_{m=1}^{N(\Delta t)} \cos^n \theta'(t_m) + \sum_{m_1=1}^{N(\Delta t)} \cdots \sum_{m_n=1}^{N(\Delta t)} \cos\theta'(t_{m_1}) \cdots \cos\theta'(t_{m_n})$$

$$m_1 \neq m_2 \neq \cdots \neq m_n .$$

The expectation over just the second term above is

$$E_{N/\Phi} \left[\sum_{m_1=1}^{N(\Delta t)} \cdots \sum_{m_n=1}^{N(\Delta t)} E_{t_m/N,\,\Phi} \{ \cos\theta'(t_{m_1}) \cdots \cos\theta'(t_{m_n}) \} \right]$$

$$m_1 \neq m_2 \neq \cdots \neq m_n$$

where $E_{t_m/N,\,\Phi}$ is a conditional expectation given N and Φ. The expectation over $N(\Delta t)$ simply becomes the average of the counting process over $(t, t + \Delta t)$. Since this expression does not involve those terms where $m_1 = m_2 = \cdots = m_n$, the above experession becomes

$$(\bar{N}^n - \bar{N})\; E_{(t_{m_1}, t_{m_2}, \cdots, t_{m_n})} [\cos\theta'(t_{m_1}) \cos\theta'(t_{m_2}) \cdots \cos\theta'(t_{m_n})]$$

where from (1-2)

$$\bar{N} = \int_{t}^{t+\Delta t} n(\tau)\, d\tau \quad .$$

The conditional expectation of the term in the brackets requires the n-dimensional joint probability density of the n random variables $\{t_m\}$. For Poisson shot noise processes this is obtained from (1-4) as

$$p(t_{m_1}, t_{m_2}, \cdots, t_{m_n} | n) = \frac{1}{\bar{N}^n} \prod_{m=1}^{n} n(t_m) \quad .$$

Therefore, the conditional expectation over the $\{t_m\}$ is

$$\int_{t}^{t+\Delta t} \cdots \int_{t+t_{m_{n-1}}}^{t+\Delta t} [\cos\theta'(t_{m_1}) \cdots \cos\theta'(t_{m_n})][n(t_{m_1}) \cdots n(t_{m_n})]\, dt_{m_1} \cdots dt_{m_n}$$

for $t \leq t_{m_1} \leq \cdots \leq t_{m_n} \leq (t + \Delta t)$. As we take the limit as Δt goes to zero this expression behaves as

$$[\cos\theta'(t_{m_1}) \cdots \cos\theta'(t_{m_n})]\, [n(t_{m_1}) \cdots n(t_{m_n})]\, (\Delta t)^n \quad .$$

Therefore, taking the limit as Δt goes to zero the above expression behaves as $(\Delta t)^{n-1}$ which goes to zero. Hence the second term resulting from Equation (2-38) is zero and

$$K_n(\Phi) = \lim_{\Delta t \to 0} \frac{(-ek)^n}{\Delta t} E_{N, t_m/\Phi} \sum_{m=1}^{N(\Delta t)} \cos^n \theta'(t_m)$$

$$= \lim_{\Delta t \to 0} \frac{(-ek)^n}{\Delta t} \bar{N} E_{t_m/\Phi} [\cos^n \theta'(t_m)] \quad .$$

The expectation of the bracketed term is

$$\int_{t}^{t+\Delta t} \cos^n \theta'(t_m) \frac{n(t_m)}{\bar{N}} dt_m$$

and

$$K_n(\Phi) = (-ek)^n \cos^n \theta'(t)\, n(t) . \tag{2-39}$$

This equation represents the general nth conditional moment of the increment of the phase error. Note that it is in terms of the feedback signal and the intensity modulation, n(t). Since Equation (2-39) is basically a product of sinusoids, $K_n(\Phi)$ will contain sine waves at the "beat" frequencies. Remembering that terms involving frequencies of $n\omega_0$, $n \geq 1$, are eliminated by the mixer*, the general expressions for $K_n(\Phi)$ become

$$K_n(\Phi) \begin{cases} -C_n (ek)^n \, \text{Absin}\, \Phi & n\text{-odd} \\ \\ C_{n-1} (ek)^n (A) & n\text{-even} \end{cases} \tag{2-40}$$

where

$$C_n = \prod_{\substack{i=1 \\ i\text{-odd}}}^{n} \left(\frac{1}{i+1} \right) \qquad n\text{-odd} \quad . \tag{2-41}$$

The series form of the pdf equation now becomes

*Mathematically, we are implying that the expectation operation in Equation (2-29) contains an additional time averaging operation, caused by the filtering effects of the mixer. Thus, to be rigorous, a time averaged version of $K_n(\Phi)$ is being computed.

$$\frac{\partial p(\Phi, t)}{\partial t} = Ab \sum_{\substack{n=1 \\ (n\text{-odd})}}^{\infty} \frac{C_n (ek)^n}{n!} \frac{\partial^n}{\partial \Phi^n} [\sin\Phi \, p(\Phi, t)] +$$

$$A \sum_{\substack{n=2 \\ (n\text{-even})}}^{\infty} \frac{C_{n-1} (ek)^n}{n!} \frac{\partial^n p(\Phi, t)}{\partial \Phi^n} . \qquad (2\text{-}42)$$

The solution to this equation is the pdf, $p(\Phi, t)$, of the phase error, Φ, at each instant of time, t. Note that the equation is an infinite order partial differential equation with coefficients that are functions of the variable Φ. The infinite number of derivative terms can be directly attributed to the "jump"nature of the phase error process. The steady state solution of the pdf is given by (2-35) obtained by setting the right side of (2-42) equal to zero. Thus, with $p(\Phi)$ denoting the steady state pdf, we have

$$0 = Ab \sum_{\substack{n=1 \\ (n\text{-odd})}}^{\infty} \frac{C_n (ek)^n}{n!} \frac{d^n}{d\Phi^n} [\sin \Phi \, p(\Phi)] + A \sum_{\substack{n=2 \\ (n\text{-even})}}^{\infty} \frac{C_{n-1} (ek)^n}{n!} \frac{d^n p(\Phi)}{d\Phi^n} . \qquad (2\text{-}44)$$

The steady state pdf can be determined by solving the above total differential equation with the approprieate initial conditions. The equation is still, however, of infinite order and the hope of obtaining an exact solution is somewhat ambitious. Nevertheless, there is still useable information that may be extracted from Equation (2-44) without a complete solution. For example, we note that the coefficients are periodic in Φ, implying that if $p(\Phi)$ is a solution to (2-44) then $p(\Phi + 2\pi)$ is also a solution. Hence, steady state pdf solutions are periodic with period 2π. For this reason we need only concentrate

on deriving a normalized solution over a single period, and Φ will therefore be constrained $(-\pi, \pi)$ in the subsequent analysis. For convenience, we can rewrite (2-44) in a slightly different form by first dividing through by the coefficient for $n = 2$. This yields

$$0 = \alpha \frac{d}{d\Phi} [\sin\Phi\, p(\Phi)] + \frac{d^2 p(\Phi)}{d\Phi^2} + \left(\frac{1}{2\alpha}\right) \frac{d^3}{d\Phi^3} [\sin\Phi\, p(\Phi)]$$

$$+ \frac{1}{4\alpha^2} \frac{d^4 p(\Phi)}{d\Phi^4} + \frac{1}{12\alpha^3} \frac{d^5}{d\Phi^5} [\sin\Phi\, p(\Phi)] \qquad (2\text{-}45)$$

where $\alpha = 2b/ek$. For a first order loop the gain k is directly related to the loop noise bandwidth B_L by [1]

$$B_L = \frac{eAk}{4} . \qquad (2\text{-}46)$$

Since it is desirous to operate the loop with a given bandwidth, the loop gain k must be adjusted to achieve this value. Hence, $k = 4B_L/eA$ and the α parameter in (2-45) takes the form

$$\alpha = \frac{Ab}{2B_L} . \qquad (2\text{-}47)$$

The coefficient Ab can be interpreted as the average rate of electrons of the intensity modulation by the synchronizing signal. In this light, α is then the average number of electrons produced in a $1/2B_L$ time period, i.e., in a time period corresponding to the reciprocal of the designed carrier bandwidth. Hence α can be considered an electron function "density", indicating the accumulation of electron occurrences over a fixed time period. By relating electron occurrences to photons, the density α can also be interpreted in terms of received synchronizing

energy, or in terms of signal to noise ratios. In particular, if we multiply numerator and denominator by e^2A, then

$$\alpha = b \frac{(eA)^2}{e^2A(2B_L)} \quad . \tag{2-48}$$

The term $(eA)^2$ is proportional to the average current power in the synchronizing signal, while (e^2A) is the spectral level of the shot noise power spectrum and $(e^2A)2B_L$ is proportional to the total shot noise power in a $2B_L$ bandwidth. Hence, α can also be considered an indication of the signal-to-shot noise power ratio. As such, we would expect performance to improve as α increases. This would mean the modulation index b should be as large as possible for best operation. We shall find this conjecture is true, and therefore from here on b will be given its maximum possible value (b = 1) in (2-47).

Note that the higher order coefficients in (2-45) decrease with increasing α. This appears to indicate a diminishing importance of the higher derivative terms in contributing to the solution as α increases. This conjecture will be investigated in the next chapter, and will be shown to have both a mathematical and physical interpretation.

One last point is worthy of comment concerning (2-45). Note that the only parameter effecting the equation, and therefore the solution, is α, the electron (photon) density in a $1/2B_L$ time period. In particular, the synchronizing carrier frequency ω_s in (2-2) does <u>not</u> appear in the solution. Hence, it is meaningless to cite values of numbers of electrons (photons) per cycle of synchronizing carrier frequency in discussing optical time locking. It is

only the number per cycle of loop bandwidth that is significant. Of course, the sync frequency is important in converting phase errors in radians to timing errors in seconds.

Chapter 3
PROBABILITY DENSITY SOLUTIONS

In Chapter 2 an infinite order differential equation was derived for the steady state probability density of the loop phase error of a first order tracking loop with shot noise inputs. The equation showed that the coefficients of the resulting derivative terms in the equation depended upon the electron function rate in the photo-detector, which in turn depended upon the received radiation power. In this chapter we investigate approximate solutions for the desired probability density of the tracking error.

3.1 High Electron Density Solution

For the case where the function density α in (2-47) is extremely high, a first approximation to the solution of Eq. (2-45) can be obtained by dropping all terms that have powers of $1/\alpha$ as coefficients. This leads to the equation

$$0 = \alpha \frac{d}{d\Phi} [\sin\Phi\, p(\Phi)] + \frac{d^2 p(\Phi)}{d\Phi^2} \tag{3-1}$$

where $p(\Phi)$ is the steady state density and α is the electron density at the photo-detector output:

$$\alpha = \frac{A}{2B_L} \quad . \tag{3-2}$$

Equation (3-1) is just the steady state form of the Fokker-Planck equation and can easily be solved. Integrating both sides yields

$$C_0 = \alpha \sin\Phi\, p(\Phi) + \frac{dp(\Phi)}{d(\Phi)} \tag{3-3}$$

where C_0 is an arbitrary constant. This equation can be solved over the interval, $-\pi \leq \Phi \leq \pi$, with the two boundary conditions:

1) $p(\pi) = p(-\pi)$ (periodicity)

2) $\int_{-\pi}^{\pi} p(\Phi)\, d\Phi = 1$.

The solution is

$$p(\Phi) = \frac{e^{\alpha \cos \Phi}}{2\pi I_0(\alpha)} \tag{3-4}$$

where I_0 is the imaginary Bessel function. Equation (3-4) is plotted in Figure 4, for various α. Note that the probability density approaches, for large α, a delta function at zero, while for $\alpha \rightarrow 0$, it approaches a uniform density over the phase error interval. The former case can be considered the limit of perfect tracking, while the latter represents a completely random phase error; i.e., poor phase tracking. The ability to track is therefore directly related to the value of the α parameter.

It is of interest to note that the solution in (3-4) is the same solution obtained for the first order loop when driven by a sinusoidal signal plus additive while Gaussian noise [1, 11]. Thus, the error differential equation due to shot noise inputs becomes identical to that due to additive input Gaussian noise as the higher order coefficients are eliminated. In essence, this serves as an apparent justification for the truncation of Eq. (2-45) to (3-1) for large values of α since it has been shown [3, 5, 6] that a discrete poisson shot noise process approaches a continuous Gaussian process as $\alpha \rightarrow \infty$. Thus, for

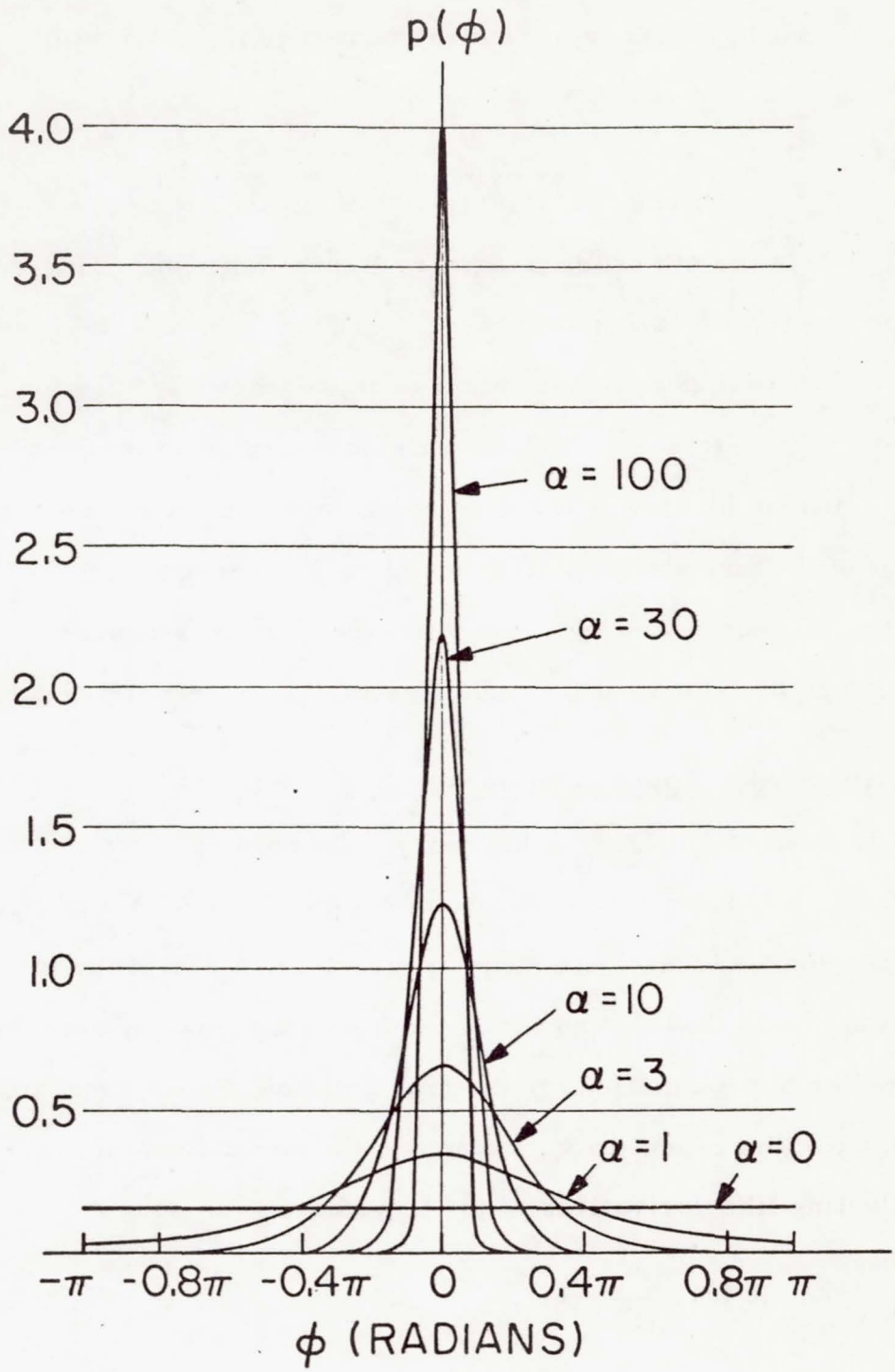

p(φ)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
α = 100
α = 30
α = 10
α = 3
α = 1
α = 0
−π
-0.8π
-0.4π
0
0.4π
0.8π
π
φ (RADIANS)

Figure 4.

$\alpha >> 1$, the shot noise error pdf is, to a first approximation, given by the solution for additive Gaussian noise inputs.

The pdf in (3-4) has zero mean and variance given by

$$\sigma_\Phi^2 = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n I_n(\alpha)}{n^2 I_0(\alpha)} \tag{3-5}$$

where $I_n(\alpha)$ is the nth order imaginary Bessel function. This variance is shown as a function of α^{-1} in Figure 5. As the parameter α approaches zero the variance approaches $\pi^2/3$, the variance of a uniformly distributed random variable over the interval, $(-\pi, \pi)$. It may be seen that the tracking variance for the steady state pdf of the phase error is approximately proportional to $1/\alpha$ for large α. For α below 5, the variance increases rapidly, but the range of validity of the high density solution is questionable.

3.2 High Order Approximations

The density in (3-4) is in theory valid only as $\alpha \to \infty$. It is not obvious, however, how accurate this solution is for finite α. In this section we investigate higher order truncations of the inifinite order equation in (2-45), and the associated solutions, in order to obtain better approximations to the true solution. After integrating (2-45) once with respect to Φ, expanding the derivatives of $\sin \Phi\, p(\Phi)$, and collecting like derivatives of $p(\Phi)$, we have

$$C_0 = \sum_{n=0}^{\infty} F_n(\Phi) \frac{d^n}{d\Phi^n} p(\Phi) \ . \tag{3-6}$$

Here C_0 is the constant of integration and the $F_n(\Phi)$ functions are of the form

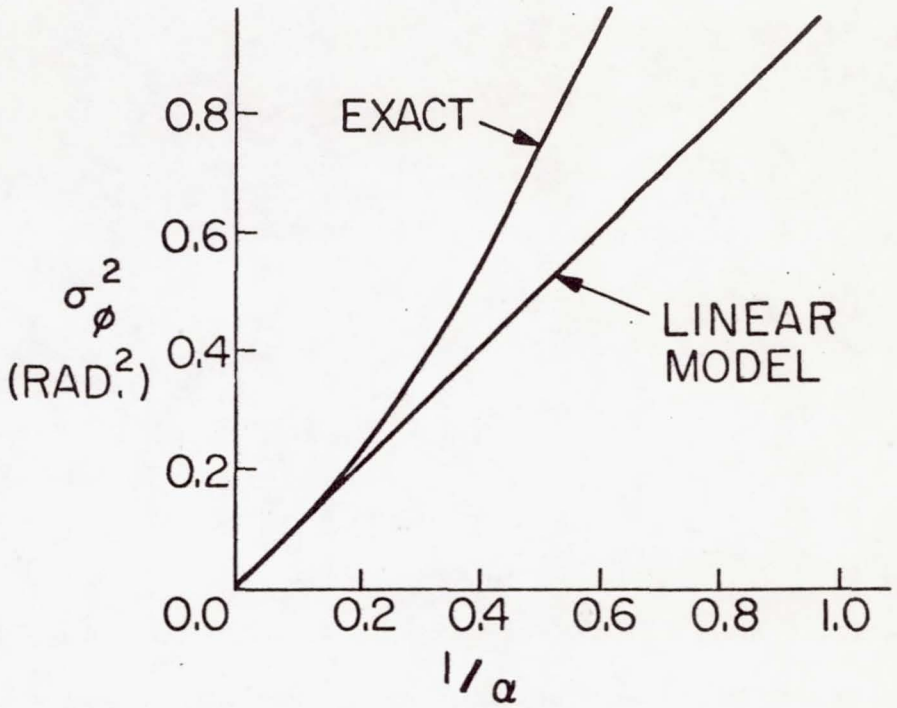

0.8
0.6
0.4
0.2
0.0
EXACT
LINEAR
MODEL
σ_ϕ^2
(RAD.^{2})
0.2
0.4
0.6
0.8
1.0
$1/\alpha$

Figure 5.

$$F_0(\Phi) = \sin\Phi\left(\alpha - \frac{1}{2\alpha} + \frac{1}{12\alpha^3} - \frac{1}{144\alpha^5} + \cdots\right)$$

$$F_1(\Phi) = 1 + \cos\Phi\left(\frac{1}{\alpha} - \frac{1}{3\alpha^3} + \frac{1}{24\alpha^5} - \cdots\right)$$

$$F_2(\Phi) = \sin\Phi\left(\frac{1}{2\alpha} - \frac{1}{2\alpha^3} + \frac{5}{48\alpha^5} - \cdots\right)$$

$$F_3(\Phi) = \frac{1}{4\alpha^2} + \cos\Phi\left(\frac{1}{3\alpha^3} - \frac{5}{36\alpha^5} + \cdots\right) \qquad (3\text{-}7)$$

$$F_4(\Phi) = \sin\Phi\left(\frac{1}{12\alpha^3} - \frac{5}{48\alpha^5} + \cdots\right)$$

$$F_5(\Phi) = \frac{1}{36\alpha^4} + \cos\Phi\left(\frac{1}{24\alpha^5} - \cdots\right)$$

$$F_6(\Phi) = \sin\Phi\left(\frac{1}{144\alpha^5} - \cdots\right)$$

etc.

Note that the functions, $F_n(\Phi)$, decrease with α (for $\alpha \geq 1$ and $n \geq 1$) and it is reasonable to assume thatsolutions to truncations involving higher order terms of Equation (3-7) may yield higher order approximations to the total solution of the finite-order differential equation. The solution to the truncated equation involving terms up to and including the jth derivative of $p(\Phi)$ will be called the jth-order truncation solution. The function, $F_n(\Phi)$, in general involves terms derived from all the odd order derivatives of order $\geq n+1$ in Equation (3-6) operating on $\sin\Phi\, p(\Phi)$. Therefore, when forming the jth truncated equation from Equation (3-6),the functions $F_n(\Phi)$ must also be appropriately truncated. For example, the solution to the Fokker-Planck equation treated previously may also be called the first-order truncation

solution to Equation (3-6). Since, for a given $\alpha \geq 1$, the functions, $F_n(\Phi)$, decrease in magnitude rapidly with n it is reasonable to expect that solutions (assuming they can be found) to increasingly higher-order truncated equations would also reduce respectively the remainder, when the higher-order truncation solutions are substituted for $p(\Phi)$ in Equation (3-6). This will be examined below as higher-order truncation solutions are found.

A method exists for solving progressively higher-order truncated versions of Equation (3-6). From Ince [14], Boyce and DiPrima [15], and Coddington and Levinson [13] it is shown that the method of Frobenius which assumes a series solution for $p(\Phi)$ of the form

$$p(\Phi) = \Phi^m \sum_{n=0}^{\infty} A_n \Phi^n , \quad A_0 \neq 0 \tag{3-8}$$

is applicable to any-order truncated version of Equation (3-6), even (in theory) the total infinite-order solution. However, to solve exactly, any nth-order truncated equation from Equation (3-6) it is necessary to have n+1 boundary conditions (recall C_0 in Equation (3-6) is an unknown constant of integration). In addition to the boundary conditions previously introduced, additional boundary conditions must be specified in order to solve the higher order differential equations.

For the non-offset case, the primary assumption that will be imposed to evaluate the necessary boundary conditions is that the solutions to (3-6) are symmetric about $\Phi = 0$. The solution is therefore an even function about $\Phi = 0$, and between $-\pi$ and π it can be expanded in a Fourier series as an infinite sum of cosines,

$$p(\Phi) = \sum_{n=0}^{\infty} a_n \cos n\Phi \qquad (3\text{-}9)$$

where the a_n's are coefficients. From this expression it can be seen that all odd order derivatives of $p(\Phi)$ are zero at $\Phi = 0$ and $\Phi = \pm\pi$. Furthermore, evaluation of the right side of (3-6) at $\Phi = 0$, with this zero condition for the odd derivatives, shows that C_0 is zero. In addition to these initial conditions, we shall further impose the restriction that all even order derivatives, evaluated at $\Phi = \pm\pi$, will be zero also. This results in the set of boundary conditions:

$$\left.\frac{d^n p(\Phi)}{d^n \Phi}\right|_{\pm\pi} = 0 \quad , \text{ for all } n \geq 1 \ . \qquad (3\text{-}10)$$

These conditions, along with the two used in (3-4) will provide a solution to any order truncation of (3-6). In the following sections, solutions to second order and third order truncated equations will be determined.

3.3 Second-Order Truncation Solution

The second-order truncation of Equation (3-6) becomes

$$\frac{\sin\Phi}{2\alpha} p''(\Phi) + \left(1 + \frac{\cos\Phi}{\alpha}\right) p'(\Phi) + \left(\alpha - \frac{1}{2\alpha}\right) \sin\Phi \; p(\Phi) = 0. \qquad (3\text{-}11)$$

The point $\Phi = 0$ is a regular singular point of Equation (3-11) and therefore by Theorem 4.3 of Boyce and DiPrima [15] a series solution exists of the form given by Equation (3-8), in either of the intervals $-\rho < \Phi < 0$ or $0 < \Phi < \rho$ where ρ is some positive number. The value of ρ is the radius of convergence of the series in Equation (3-8), and is at least equal to the distance from the origin to the nearest zero of $\sin\Phi/2\alpha$, which is at π. Hence, a series solution can be found for Φ

in the range $-\pi$ to π for which the series converges.

By writing $\sin\Phi$ and $\cos\Phi$ in their series expansions, substituting Equation (3-8) into Equation (3-11), and collecting like powers of Φ, solutions for m and A_n can be found. Two solutions are found for m, one being zero and the other nonzero. Only the zero value for m yields a non-trivial results and the resulting values for A_n, n even, are

$$A_n = \frac{(-1)^{\frac{n+2}{2}} \sum_{\substack{r=0 \\ (r\text{-even})}}^{n-2} (-1)^{\frac{r}{2}} \left[\frac{(r-1)r}{(n+1-r)!} + \frac{2r}{(n-r)!} - \frac{\beta}{(n-1-r)!} \right] A_r}{n(\lambda + 1 + n)} \qquad (3\text{-}12)$$

where $\beta = 2\alpha^2 - 1$, $\lambda = 2\alpha$. A_n, for n odd are all zero since the density is symmetrical. Therefore, for given values of α, all the necessary coefficients, A_n, can be calculated to solve for $p(\Phi)$ in its series expansion. This was carried out on a digital computer for α equal to 1.5, 3, 10, and 30. The right half of the symmetrical density $p(\Phi)$ in (3-8) is plotted in Figures 6, 7, 8, and 9 for these values, along with the solutions to the Fokker-Planck equation for the same α. Note that the truncated solution converges rather quickly to the high density solution, and are practically equivalent for $\alpha \geq 3$. In essence, this can be conjectured as the range of validity of the high density solution. The variance of the phase error, calculated from (3-8), is also shown in Figure 10, along with the variance of the high density solution, Equation (3-5), and that satisfying a linear relation in $1/\alpha$. Again, the results indicate that for $\alpha \geq 2$, the relation in (3-5) is valid for the second order truncation solution as well.

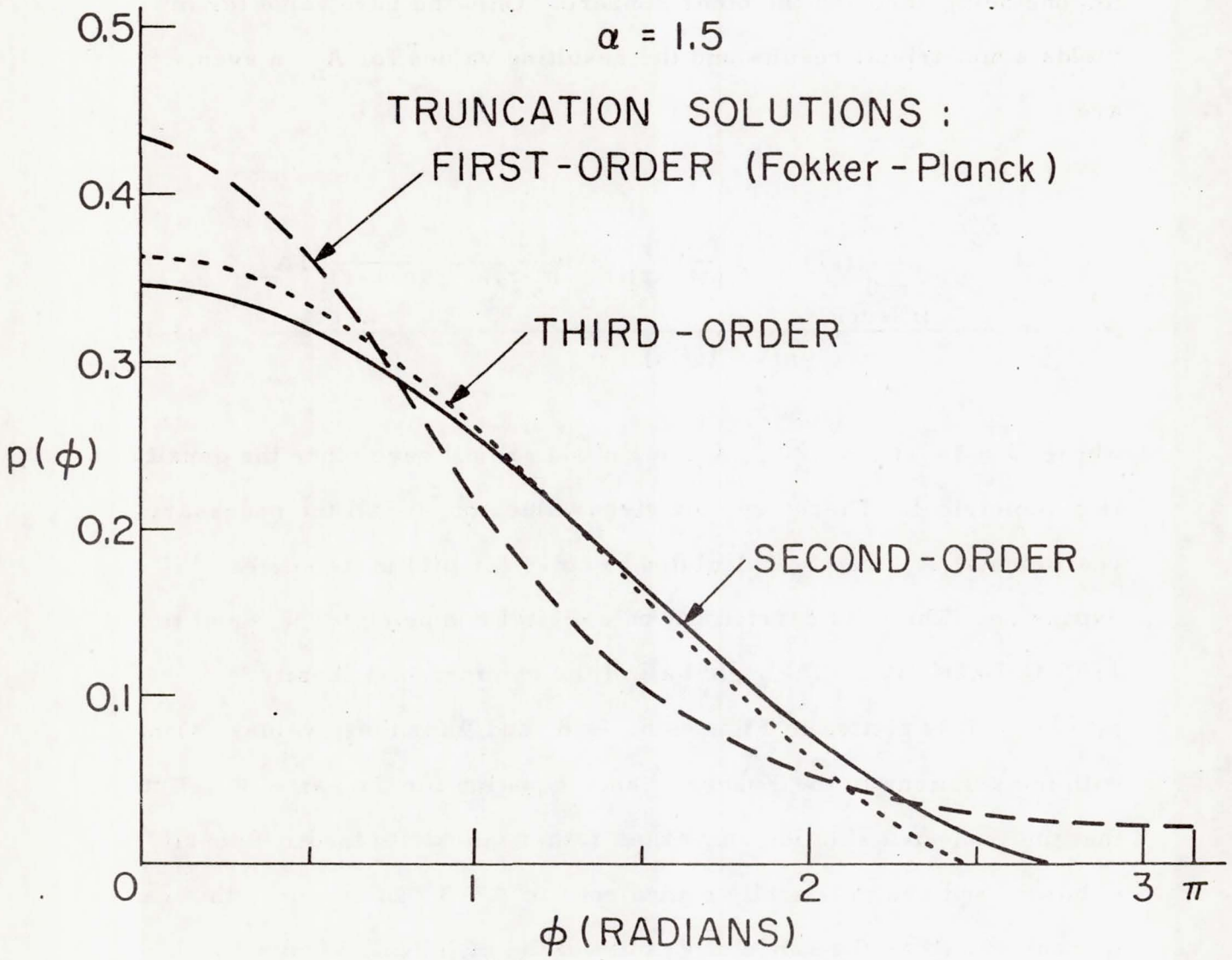

α = 1.5
TRUNCATION SOLUTIONS :
FIRST-ORDER (Fokker-Planck)
THIRD-ORDER
SECOND-ORDER
p(φ)
0.5
0.4
0.3
0.2
0.1
0
1
2
3 π
φ (RADIANS)

Figure 6.

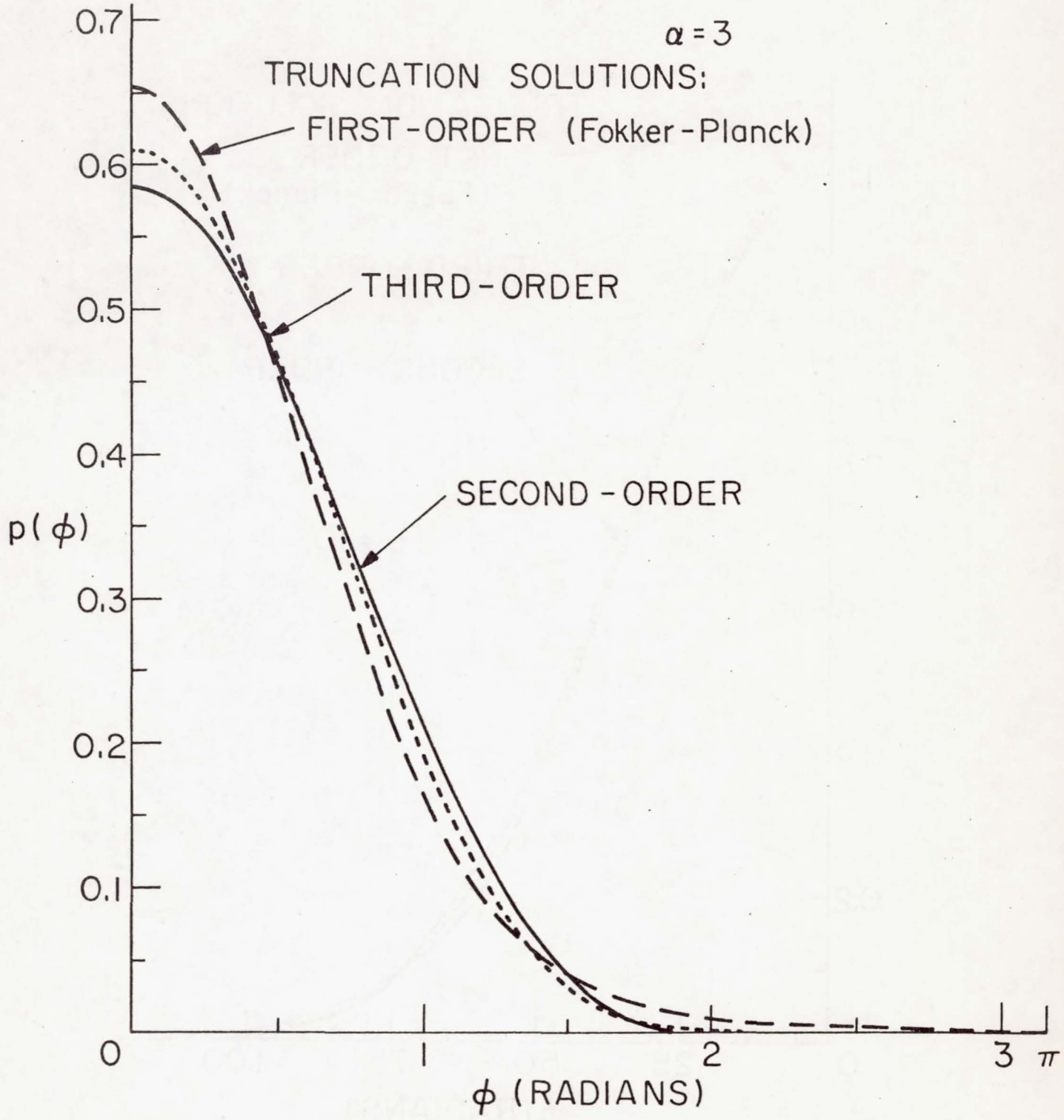

α = 3
TRUNCATION SOLUTIONS:
FIRST-ORDER (Fokker-Planck)
THIRD-ORDER
SECOND-ORDER
p(φ)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3 π
φ (RADIANS)

Figure 7.

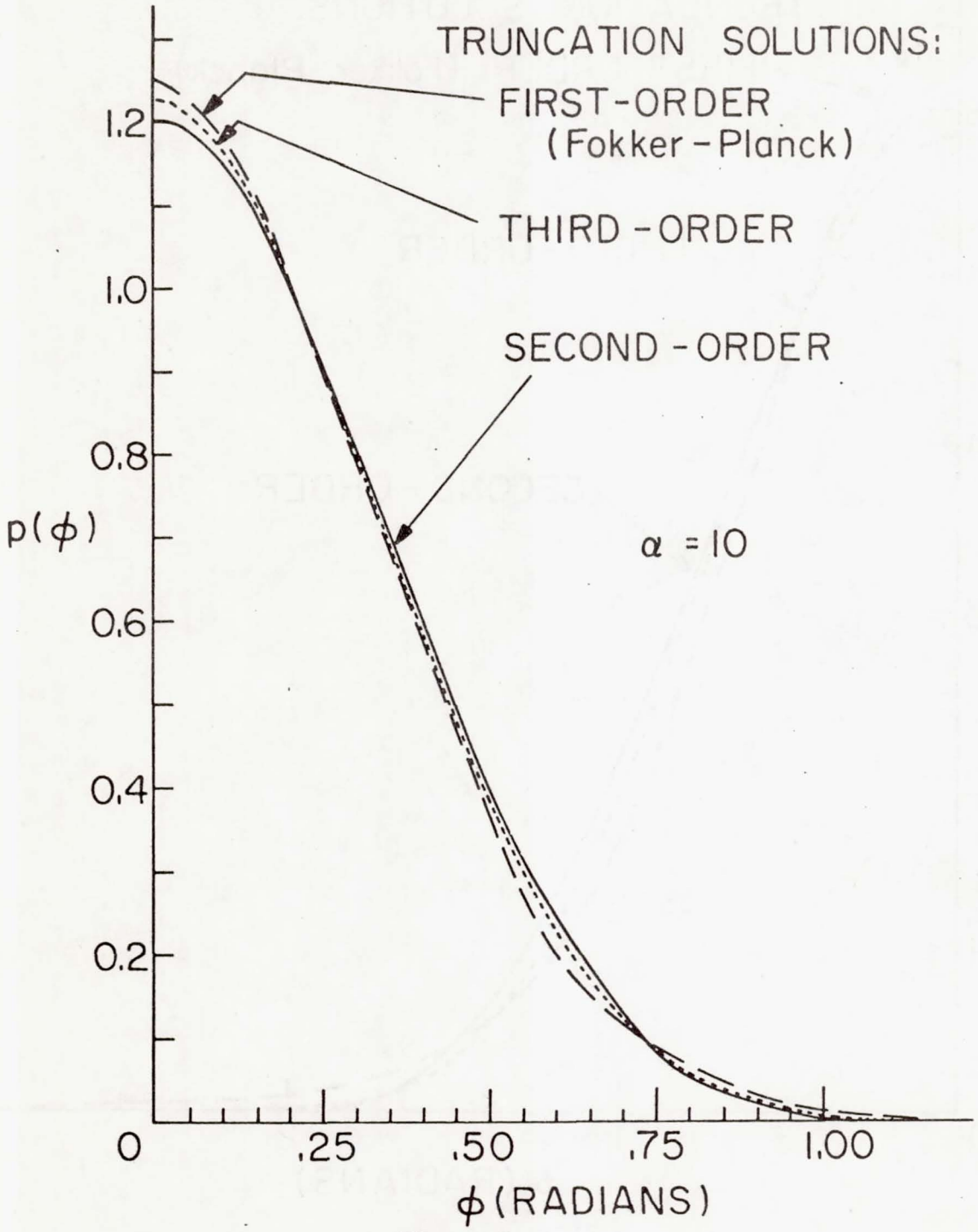

TRUNCATION SOLUTIONS:
FIRST-ORDER
(Fokker-Planck)
THIRD-ORDER
SECOND-ORDER
α = 10
p(φ)
1.2
1.0
0.8
0.6
0.4
0.2
0
.25
.50
.75
1.00
φ (RADIANS)

Figure 8.

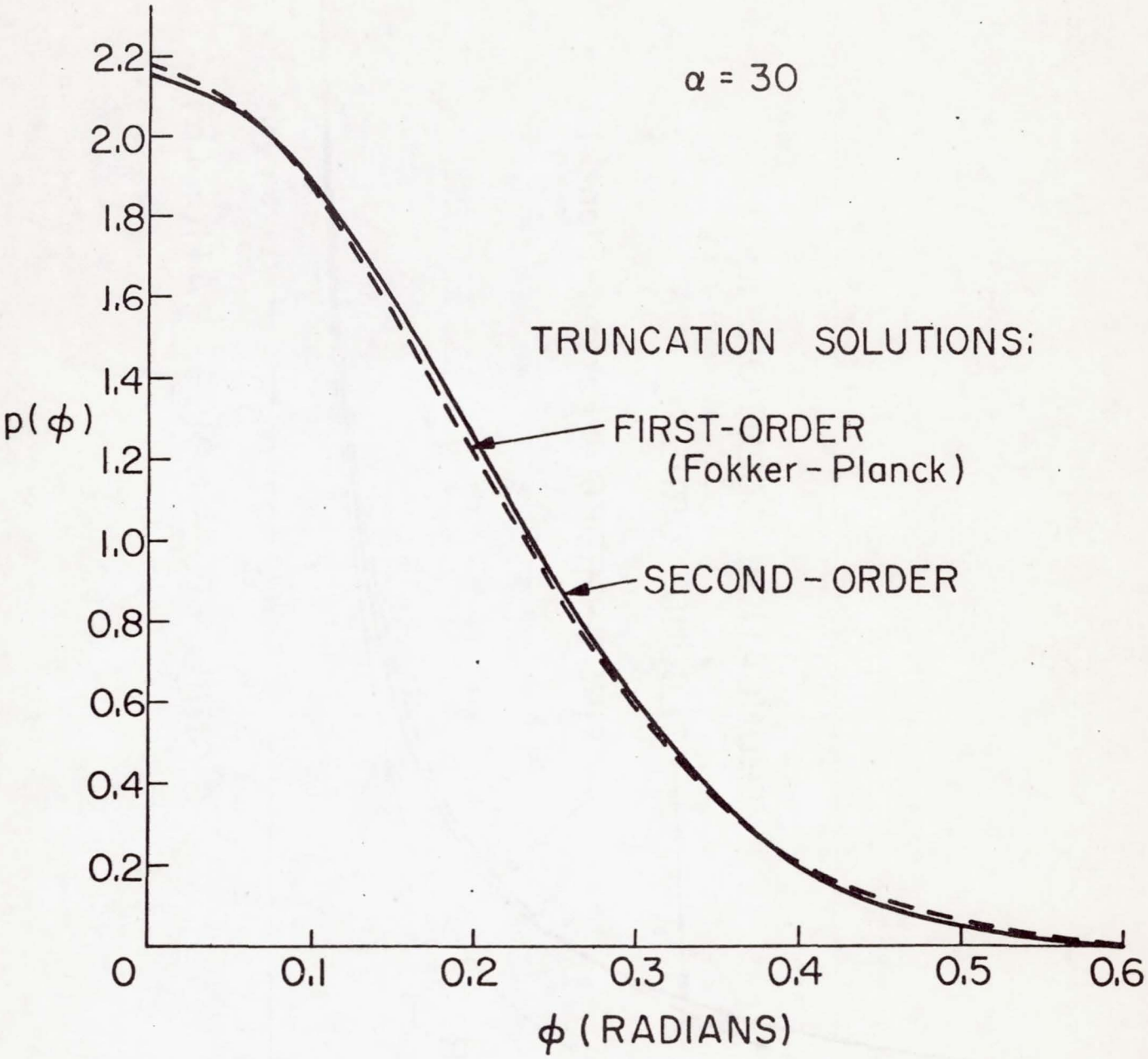

α = 30
TRUNCATION SOLUTIONS:
FIRST-ORDER
(Fokker - Planck)
SECOND - ORDER
p(φ)
φ (RADIANS)
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6

Figure 9.

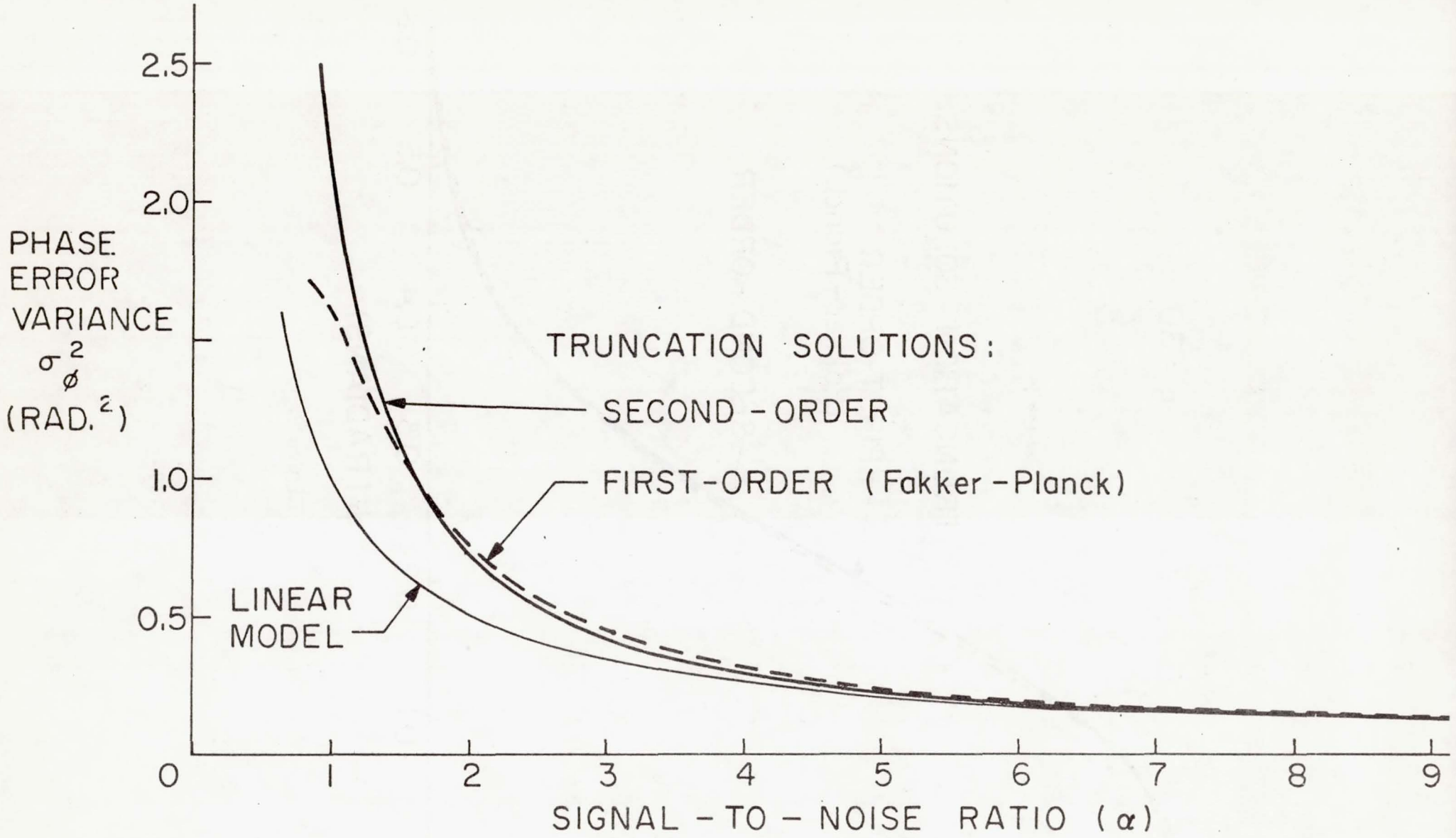

2.5
2.0
1.0
0.5
0
PHASE ERROR VARIANCE σ_{ϕ}^2 (RAD.^{2})
TRUNCATION SOLUTIONS:
SECOND - ORDER
FIRST-ORDER (Fakker - Planck)
LINEAR MODEL
1
2
3
4
5
6
7
8
9
SIGNAL - TO - NOISE RATIO (α)

Figure 10.

3.4 Third-Order Truncation Solution

The third-order truncation of Equation (3-6) gives

$$\frac{1}{4\alpha^2} p'''(\Phi) + \frac{\sin\Phi}{2\alpha} p''(\Phi) + \left(1 + \frac{\cos\Phi}{\alpha}\right) p'(\Phi) + \left(\alpha - \frac{1}{2\alpha}\right) \sin\Phi \; p(\Phi) = 0 . \tag{3-13}$$

To solve this equation the previous series method is also used. However, $\Phi = 0$ is no longer a singular point for this equation, and the series solution is simplified slightly to

$$p(\Phi) = \sum_{n=0}^{\infty} A_n \Phi^n \qquad A_0 \neq 0 \quad . \tag{3-14}$$

The four boundary conditions used here are

i) $p(\pi) = p(-\pi)$

ii) $\int_{-\pi}^{\pi} p(\Phi)\, d\Phi = 1$

iii) $p'(\Phi)\big|_{\pi} = 0$

iv) $p''(\Phi)\big|_{\pi} = 0.$

Boundary conditions i) and iii) imply all A_n (n-odd) are equal to zero. Use of the same method to determine A_n (n-even) as was used previously, yields the recurrence relation

$$A_n = \frac{\lambda}{n(n-1)(n-2)} \left\{ -\left[(n-2)(n+\lambda) \right] A_{n-2} + (-1)^{\frac{n+2}{2}} \sum_{\substack{r=0 \\ (r\text{-even})}}^{n-4} (-1)^{\frac{r}{2}} \left[\frac{-r(r-1)}{(n-r-1)!} - \frac{2r}{(n-r-1)!} + \frac{\beta}{(n-r-3)!} \right] A_r \right\} \tag{3-15}$$

for $n \geq 4$ and $\lambda = 2\alpha$, $\beta = 2\alpha^2 - 1$. Boundary condition iv) is used to

determine A_2. Substitution yields

$$\sum_{n=2}^{\infty} n(n-1)\, A_n\, \pi^{n-2} = 0 \,. \tag{3-16}$$

Since all the $A_n (n \geq 2)$ can be written in terms of A_0 and A_2 Equation (3-16) can be written as

$$\left[\sum_{n=2}^{\infty} n(n-1)\, \pi^{n-2}\, D_n \right] A_2 + \left[\sum_{n=4}^{N} n(n-1)\, \pi^{n-2}\, B_n \right] A_0 = 0$$

where D_n and B_n can be determined from Equation (3-15). Then A_2 is

$$A_2 = - \frac{\left[\sum_{n=4}^{N} n(n-1)\pi^{n-2}\, B_n \right]}{\left[\sum^{N} n(n-1)\pi^{n-2}\, D_n \right]} \; A_0 \quad .$$

A_0 is then determined by the normalizing boundary condition (ii),

$$\int_{-\pi}^{\pi} p(\Phi) = 1 \quad .$$

These computations were also accomplished with a digital computer and the solutions for $p(\Phi)$, $0 \leq \Phi \leq \pi$, are plotted in Figures 6, 7, and 8 for $\alpha = 1.5$, 3, and 10, respectively. For $\alpha \geq 1.5$, the third order truncation solution is almost identical to the second order solution.

3.5 Accuracy of the Truncation Solutions

The preceeding methods can be used to solve higher-order truncations of Equation (3-6) but the derivation of the expressions for A_n become increasingly more difficult and computer time and size (memory) required increase quite rapidly. Therefore truncation solutions of order higher than three were not attempted. However, it would be of interest to obtain an indication of how well the truncation solutions were approximating the true solution to (2-45). In particular, it is desirable to justify the notion that each succeedingly higher-order truncation solution was a better approximation to the total solution. This requires that the truncation solutions be substituted into Equation (3-6), and the magnitude of the remainder associated with the higher order neglected terms should be investigated.

With this objective the solutions obtained for the first-order (Fokker-Planck) and second-order truncations were substituted into Equation (3-6) and the magnitude of the maximum value of the remaining terms were calculated on a computer. The results are plotted in Figure 11 for various values of α. For example, when the first-order solution was used, the largest remainder was due to the second-order term, the next largest due to the third-order term, etc. In addition, the magnitude of the third-order term, when the second-order solution is used, is smaller than it was when the first-order solution was used. It is clear from studying Figure 11 that succeedingly higher-order truncation solutions result in smaller remainders, and therefore provide a more accurate approximation to the total solution. Note also that while Figure 11 plots the maximum magnitude of each term, the sign of the remainder terms alternate. Hence, the remainder appears as an

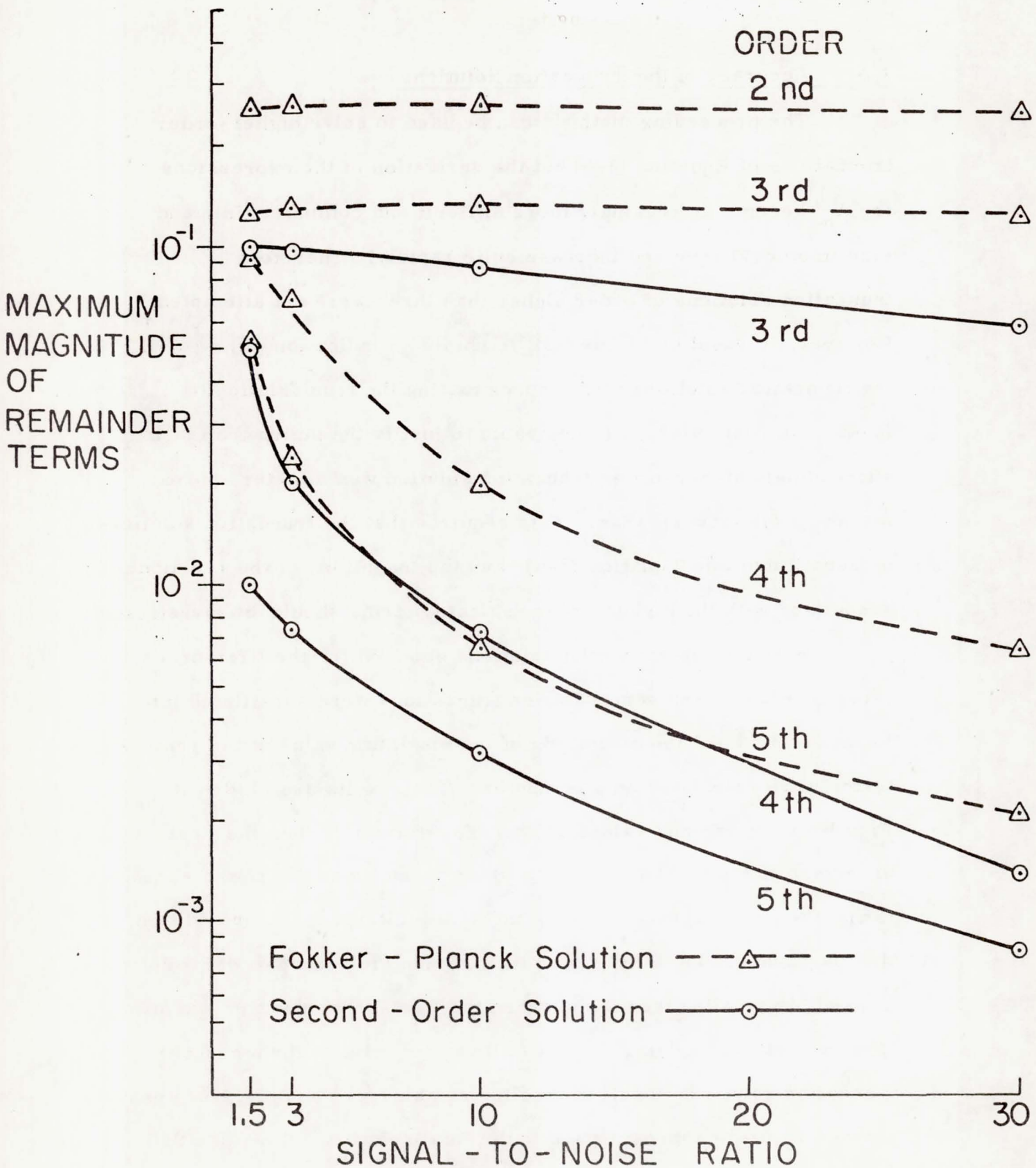

ORDER
2 nd
3 rd
10^{-1}
MAXIMUM MAGNITUDE OF REMAINDER TERMS
3 rd
4 th
10^{-2}
5 th
4 th
5 th
10^{-3}
Fokker - Planck Solution
Second - Order Solution
1.5
3
10
20
30
SIGNAL - TO - NOISE RATIO

Figure 11.

alternating series of decreasing terms and the magnitude of the remainder for an nth-order solution is bounded by the magnitude of the n+1 remainder term. For example, for $\alpha = 10$, the remainder for the Fokker-Planck solution is bounded by the second-order value of 0.26, and the remainder for the second-order solution is bounded by the third-order value of 0.086, down 67%.

From the data presented in Figures 6 through 11 it is indicative that for the low function density case higher-order truncation solutions to Equation (3-6) yield better approximations to the total solution of the infinite-order equation. It is also quite clear that as α increases all the truncation solutions approach the first-order (Fokker-Planck) solution. In other words, the nth-order truncation solution may be represented by

$$p(\Phi) = p_1(\Phi) + p_n^*(\Phi) \tag{3-17}$$

where $p_1(\Phi)$ is the solution to the first-order (Fokker-Planck) equation in (3-4) and $p_n^*(\Phi, \alpha)$ represents the difference between the nth-order and first-order truncation. As α gets very large

$$\lim_{\alpha \to \infty} p_n^*(\Phi) \to 0 \qquad n > 1$$

and

$$\lim_{\alpha \to \infty} p_n(\Phi) \to p_1(\Phi) \qquad n > 1 \quad .$$

The method of solution that has been presented here can reduce this error to as small a number as desired, in theory, given enough time and computer capacity. The third-order truncation solution was the highest-order computed in this analysis and it is shown that this solution is a good compromise in the tradeoff between accuracy and complecity of solution for the range $\alpha \geq 1.5$.

3.6 The VCO Offset Case

It has been assumed that the carrier frequency of the optical modulating signal, ω_s, and the phase-locked-loop VCO rest frequency, ω_0, have been equal. When this is not the case the VCO offset, $(\omega_s - \omega_0)$, must be included in (2-37). This expression for $\Delta\Phi$ is then modified to

$$\Delta\Phi = (\omega - \omega_0)\,\Delta t - ek \sum_{m=1}^{N(\Delta t)} \cos\,\theta'(t_m)$$

$$t \leq t_m \leq t + \Delta t \quad . \tag{3-18}$$

The $K_n(\Phi)$ coefficients in the Smoluchowski series equation are modified only through the first one which becomes

$$K_1(\Phi) = (\omega - \omega_0) - \frac{1}{2}\, eAK \sin\Phi \ .$$

The effect of the VCO offset is such that $p(\Phi)$ is no longer symmetrical. This means that the series method of finding solutions to truncations of the infinite-order Smoluchowski equation now has the odd as well as the even terms in the power series solution for $p(\Phi)$. In addition, the constant C_0 is no longer zero.

As an example of the treatment of the VCO offset case a second-order truncation solution will be found. The pertinent equation is a modified version of (3-11),

$$\frac{\sin\Phi}{2\alpha}\, p''(\Phi) + \left(1 + \frac{\cos\Phi}{\alpha}\right) p'(\Phi) + \left(\alpha - \frac{1}{2\alpha} - \gamma\right) \sin\Phi\, p(\Phi) = C_0 \tag{3-19}$$

where

$$\gamma = \frac{8(\omega - \omega_0)}{k^2(2ne^2A)} \tag{3-20}$$

is the new parameter due to the offset. For a specific value of γ, (3-19) can be solved by the series method of the previous sections. For example, with $\gamma = (.707)^\alpha$, the A_n coefficients in the series solution become

$$A_1 = \frac{\lambda C_0 + .3535\lambda^2 A_0}{\lambda + 2}$$

$$A_{n\,(n\text{-even})} = \frac{.3535\lambda^2 A_{n-1} + (-1)^{\frac{n+2}{2}} \sum_{\substack{r=0 \\ (r\text{-even})}}^{n-2} (-1)^{\frac{r}{2}} \left[\frac{(r-1)r}{(n-r+1)!} + \frac{2r}{(n-r)!} - \frac{\beta}{(n-r-1)!} \right] A_r}{n(n+\lambda+1)}$$

$$A_{n\,(n\text{-odd})} = \frac{.3535\lambda^2 A_{n-1} (-1)^{\frac{n+3}{2}} \sum_{\substack{r=1 \\ (r\text{-odd})}}^{n-2} (-1)^{\left(\frac{r+1}{2}\right)} \left[\frac{(r-1)r}{(n-r+1)!} + \frac{2r}{(n-r)!} - \frac{\beta}{(n-r-1)!} \right] A_r}{n(n+\lambda+1)}$$

where $\lambda = 2\alpha$ and $\beta = 2\alpha^2 - 1$. The two unknown constants, C_0 and A_0, can be evaluated by using the two boundary conditions

i) $p(\pi) = p(-\pi)$

ii) $\int_{-\pi}^{\pi} p(\Phi)\, d\Phi = 1.$

This was accomplished on a digital computer for $\alpha = 1.5$ and 3 and the results are plotted in Figure 12 along with the first-order solution for $\alpha = 3$. The obvious difference between this case and the non-offset

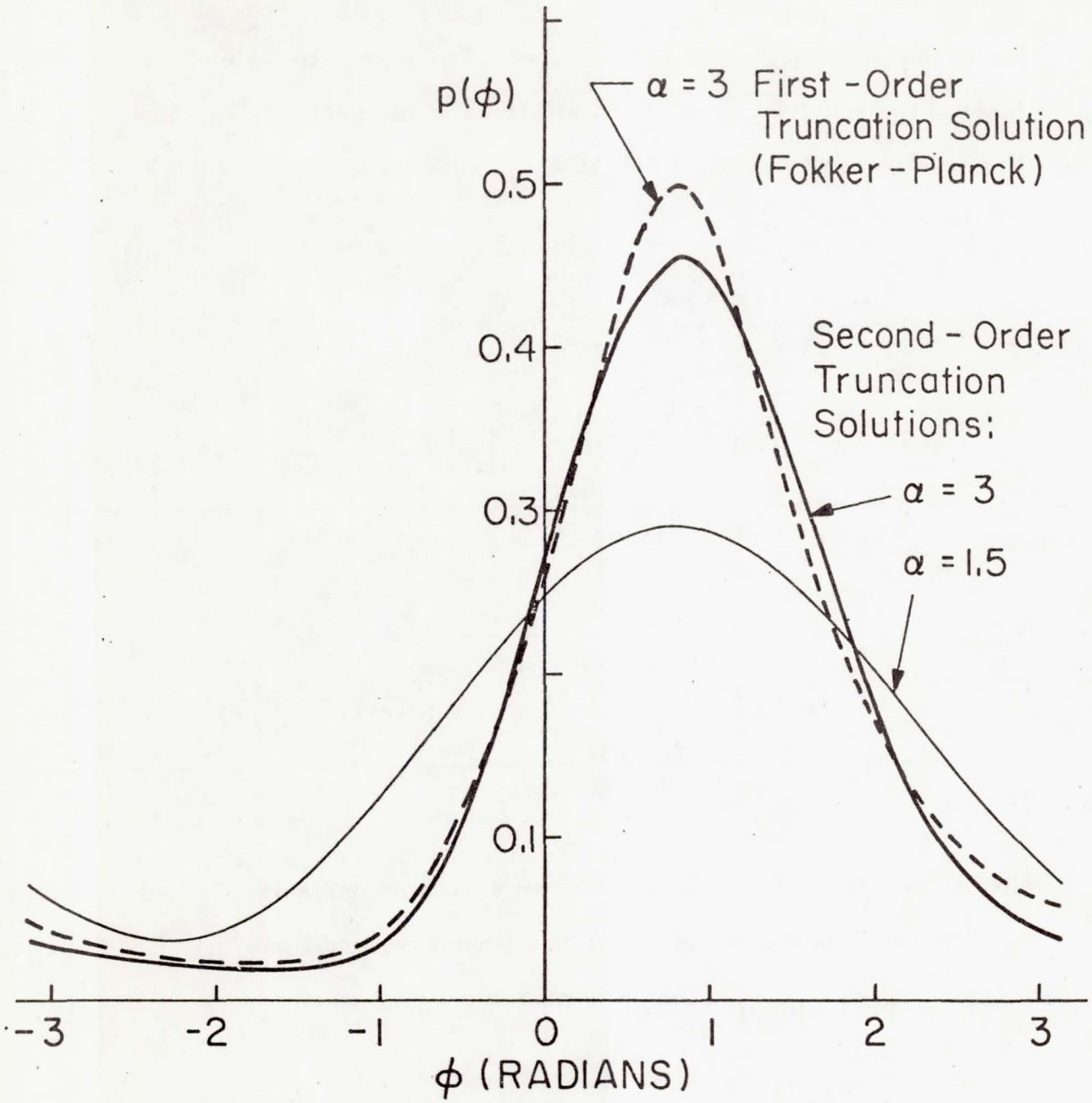

p(φ)
α = 3 First - Order
Truncation Solution
(Fokker - Planck)
0.5
0.4
Second - Order
Truncation
Solutions:
α = 3
0.3
α = 1.5
0.1
-3
-2
-1
0
1
2
3
φ (RADIANS)

Figure 12.

case is that the peak of the probability density has now shifted from the $\Phi = 0$ center line.

The two solutions for $\alpha = 3$ show approximately the same relationship as in Figure 8 for the non-offset case.

Higher-order solutions can also be obtained as in previous sections for the non-offset cases if additional boundary conditions are imposed to evaluate all the unknown constants of integration. The equivalent order offset solution, however, is obtained with more difficulty and complexity than in the non-offset case because C_0 is no longer zero.

Chapter 4

THERMAL NOISE AND PHOTOMULTIPLIERS EFFECTS

In the previous chapters a first order phase locked loop driven by a shot noise process was considered. In this chapter we investigate the effects of additive thermal noise and photomultiplier devices preceeding the loop.

4.1 Additive Gaussian Thermal Noise

Let r(t) represent a zero mean stationary Gaussian noise process having a flat one-sided power spectral density of N_0 watts/hz. When r(t) is added to the shot noise input process of the phase lock loop of Figure (1-2), the output of the loop filter [previously (2-6)] is now

$$e(t) = k_1 \int_0^t f(t-\tau)\left[\sum_{m=1}^{N(0,\tau)} e\,\delta(\tau - \tau_m)\cos\theta'(\tau) + r'(\tau)\right]d\tau \qquad (4\text{-}1)$$

where r'(t) is the "low frequency" equivalent noise process obtained by mixing the input noise r(t) with the VCO process. It has been shown [1] that the new noise term is itself Gaussian, zero mean, with spectral density given by N_0 ; (i.e., r'(t) is simply a "frequency shifted" version of r(t)).

When the transmitted phase variation, $\theta_1(t)$, is a constant, the phase error derivative for the first-order loop has the form

$$\frac{d\Phi}{dt} = -ke\left[\sum_{m=1}^{N(t)} \delta(t-t_m)\cos\theta'(t) + r'(t)\right] . \qquad (4\text{-}2)$$

If this equation is integrated from t to t + Δt, the incremental phase error becomes

$$\Delta\Phi = ek \sum_{m=1}^{N(\Delta t)} \cos\theta'(t_m) - k \int_{t}^{t+\Delta t} r'(\tau)\, d\tau \quad . \tag{4-3}$$

The first term is identical to that previously derived in (2-37). The second term accounts for the added effect of the thermal noise. The coefficients of the Smoluchowski equation can now be recalculated for $\Delta\Phi$ of Equation (4-3). In particular, $K_1(\Phi)$ remains the same as before:

$$K_1(\Phi) = - e \frac{Ak}{2} \sin\Phi \tag{4-4}$$

since the expected value of the Gaussian process is zero. The second moment requires calculation of

$$E\left[-ek \sum_{m=1}^{N(\Delta t)} \cos\theta'(t_m) - k \int_{t}^{t+\Delta t} r'(\tau)\, d\tau \right] . \tag{4-5}$$

The expectation of the square of the first term has previously been calculated, the expectation of the cross term is zero, and the expectation of the square of the second term is $k^2 N_0/2$. Therefore

$$K_2(\Phi) = \frac{k^2}{2} \left[e^2 A + N_0 \right] . \tag{4-6}$$

For computing the higher amounts, $A_n(\Phi)$, define

$$P = e \sum_{m=1}^{N(\Delta t)} \cos\theta'(t_m)$$

$$G = \int_{t}^{t+\Delta t} r'(\tau)\, d\tau .$$

Then,

$$E[P + G]^n = E(P^n + a_{n-1} P^{n-1} G + a_{n-2} P^{n-2} G^2 + a_{n-3} P^{n-3} G^3 + \cdots) .$$

Since the Poisson and Gaussian processes are independent, this becomes

$$E[P+G]^n = [EP^n + a_{n-1} E(P^{n-1}) E(G) + a_{n-2} E(P^{n-2}) E(G^2) + \cdots] .$$

It has been shown [1] that for

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E(G^n) = 0 \qquad n > 2 .$$

The expectation of P^{n-m} has already been calculated in (2-33) and been found proportional to Δt. Thus

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} a_{n-2} E(P^{n-m}) E(G^m) = 0 \qquad \begin{matrix} n \geq 2 \\ m \geq 1 \end{matrix}$$

and therefore,

$$K_n(\Phi) = \lim_{\Delta t \to 0} \frac{(-k)^n}{\Delta t} E(P^n) \qquad n > 2 \qquad (4-7)$$

which is the same as in the earlier section when no additive Gaussian noise was present. Hence, the Smoluchowski series equation has been modified only in the second term, $K_2(\Phi)$. The solution for the probability density of the phase error again requires solution of (2-45) with the appropriate K_2 modification. It has already been shown that an excellent approximate solution for the high α case is the solution to the Fokker-Planck equation. For the new $K_2(\Phi)$ term this becomes

$$P(\Phi) \tilde{=} \frac{e^{\alpha \cos \Phi}}{2\pi I_0(\alpha)} \qquad (4-8)$$

with the parameter α is redefined as

$$\alpha = \frac{\frac{1}{2}(eA)^2}{B_L\left[e^2A + N_0\right]} \quad . \tag{4-9}$$

Note that the parameter α now takes on a slightly different meaning. The bracketed denominator term is the sum of the spectral level due to the shot noise and the spectral level of the additive Gaussian noise. Hence, the denominator represents the total effective noise in the $2B_L$ loop bandwidth, due to both the shot noise and additive noise. The numerator is the average power of the intensity process. Thus, α now plays the role of an operating signal-to-noise power ratio in the tracking loop bandwidth. The dependence of $p(\Phi)$ in (4-8) on α had been shown earlier in Figure 4, and the results there are valid with above interpretation of α.

Effect of Photomultiplication

In many optical systems photomultiplication is used at the photodetector to enhance the received signal. The objective of this section is to investigate the effects of photomultiplication on the behavior of the phase error in a first-order tracking system.

An ideal photomultiplication of gain G has the property that it produces G electrons at the photodetector output for each photo-electron at the input. If the electrons are considered identical this has the effect of producing an equivalent electron pulse waveform whose magnitude is G times the magnitude of a single electron pulse waveform. Effectively, this increases the charge of a single electron by the gain G. The shot noise current of Equation (2-1) may then be written as

$$x(t) = \sum_{m=1}^{N(t)} (eG)\delta(t-t_m) . \qquad (4-10)$$

The pdf for the phase error in the high function density case is again given by (3-4), where the signal to noise ratio parameter α is now

$$\alpha = \frac{\frac{1}{2}(eGA)^2}{B_L[(eG)^2 A + N_0]} \qquad (4-11)$$

and B_L is now eGAk/4. The photomultiplication advantage is easily seen when the additive noise term of power spectrum level N_0 is dominant. In this case an increase in the α parameter can be achieved by increasing the photomultiplication gain G.

In the practical fabrication of photomultipliers the gain itself is often a random variable. In the following it is assumed that the photomultiplier has a statistically variable gain which is a random variable with mean $\overline{G}$ and mean square $\overline{G^2}$. This means that each electron at the input produces G electrons at the output, where G is a positive random variable. The shot noise current now becomes

$$x(t) = \sum_{m=1}^{N(t)} eG_m \delta(t-t_m) \qquad (4-12)$$

where the $\{G_m\}$ constitutes a set of random variables, independent, and identically distributed over zero to infinity. The incremental change in the process is now

$$\Delta\Phi = -ek \sum_{m=1}^{N(\Delta t)} G_m \cos\theta'(t_m) - k \int_{t}^{t+\Delta t} r'(\tau)\, d\tau$$

and the first two moments become

$$K_1(\Phi) = - \frac{ek\,\bar{G}\,A}{2} \sin\Phi$$

$$K_2(\Phi) = \frac{k^2}{2}\left[(ek)^2\,\overline{G^2}(A) + N_0\right] .$$

The signal-to-noise ratio α is modified to

$$\alpha = \frac{\frac{1}{2}(e\bar{G}A)^2}{B_L\left[e^2\,\overline{G^2}\,A + N_0\right]} \qquad (4\text{-}12)$$

with

$$B_L = e\,\frac{\bar{G}AK}{4} .$$

Hence, the shot noise power spectrum is increased by the mean square of the gain, while the signal power is increased by the square of the mean gain.

In some analyses it is common to assume the random gain is Gaussian with a mean $\bar{G}$, and a standard deviation, or "spread", given as a fraction of the mean gain. That is,

$$\sigma_G = \rho\frac{\bar{G}}{2} \qquad 0 \leq \rho \leq 1 .$$

In this case the mean square gain is

$$\overline{G^2} = (\bar{G})^2\left(1 + \frac{\rho^2}{4}\right) \quad 0 \leq \rho \leq 1$$

and Equation (2-41) becomes

$$\alpha = \frac{\frac{1}{2}(e\bar{G}A)^2}{B_L\left[e^2(\bar{G})^2(1+\frac{\rho^2}{4})A + N_0\right]} \quad .$$

Note that the α parameter degrades as the "spread" parameter ρ increases.

Chapter 5
SECOND-ORDER LOOP ANALYSIS AND THE GENERAL TRACKING LOOPS

In this chapter the analysis of phase-locked-loops with shot noise inputs is extended to loops of order greater than one, and to the first-order loop where the input and feedback functions are not necessarily sinusoidal (general tracking loops). For the second-order loop a vector form of the Smoluchowski equation is used for the phase error probability density, and solution can be approximated under conditions similar to those of the first-order loop. For the general tracking loop, a generalized Smoluchowski equation for the probability density is used, and again can be solved by the numerical techniques presented in Chapter 3.

5.1 The Two Dimensional Smoluchowski Equation

The Smoluchowski-Kolmogorov probability density equation was derived in Chapter 2 for a general scalar random process $\Phi(t)$. The same basic procedure can be repeated for a vector random process, and a similar vector form of (2-28) will result. Specifically, if we denote

$$\underline{\Phi}(t) = \{\Phi_1(t),\ \Phi_2(t)\} \tag{5-1}$$

as the two dimensional vector process having scalar random component processes $\{\Phi_i(t)\}$, then the vector equivalent of (2-18) is

$$P(\underline{\Phi}_1, t_1) = \int P(\underline{\Phi}_1,\ t_1 \mid \underline{\Phi}_2,\ t_2)\, P(\underline{\Phi}_2, t_2)\, d\underline{\Phi}_2 \tag{5-2}$$

where $\underline{\Phi}_i = \{\Phi_1(t_i),\ \Phi_2(t_i)\}$. Defining the two dimensional equivalent

of the characteristic function in (2-19), and repeating the steps in (2-20) through (2-28) will yield the equation

$$\frac{\partial P(\underline{\Phi}, t)}{\partial t} = \sum_m \sum_n \frac{(-1)^{m+n}}{m!n!} \frac{\partial^{m+n}}{\partial \Phi_1^m \partial \Phi_2^n} [K_{mn}(\underline{\Phi}) P(\underline{\theta}, t)] \tag{5-3}$$

where

$$K_{mn}(\underline{\Phi}) = \lim_{\Delta t \to 0} \frac{E[(\Delta \Phi_1)^m (\Delta \Phi_2)]}{\Delta t} \quad . \tag{5-4}$$

Equation (5-3) is just the two dimensional equivalent of the Smoluchowski equation in (2-28). Note that evaluation of the coefficients $\{ K_{mn} \}$ require all the statistical cross-moments of the joint variations $\Delta \Phi_1$ and $\Delta \Phi_2$ in the components of the process $\underline{\Phi}$ (t). In the following section we apply (5-3) to a second order phase lock tracking loop.

5.2 Second Order Phase Lock Loops

A second order phase lock loop is one in which the loop filter in Figure 2 introduces an integration. The basic form of such a filter would be one having transfer function

$$F(s) = \frac{s+a}{s} = 1 + \frac{a}{s} \tag{5-5}$$

where a represents a possible zero of transmission. The impulse response corresponding to (5-5) is then

$$f(t) = \delta(t) + a \quad .$$

For a shot noise input, the general loop error dynamical equation in (2-9) now becomes

$$\frac{d\Phi(t)}{dt} = -ek\cos\theta'(t)\sum_{m=1}^{N(t)}\delta(t-t_m) - aek\sum_{m=1}^{N(t)}\cos\theta'(t_m)\,. \qquad (5\text{-}6)$$

We now see that the variation $\Delta\Phi$ will contain a term in Φ [as in (2-37)] and a term involving an integral in Φ, due to the second term in (5-6). It is therefore convenient to define the vector

$$\underline{y}(t) = \{y_0(t),\ y_1(t)\} \qquad (5\text{-}7)$$

where

$$y_1(t) = \frac{dy_0(t)}{dt} \qquad (5\text{-}8)$$

and let

$$\Phi(\tau) = y_0(t) + y_1(t)\ . \qquad (5\text{-}9)$$

That is, we consider the error process in (5-6) to be decomposed into the sum of the components of a vector process $\underline{y}(t)$. The probability density of $\Phi(t)$ is then determined from the joint probability density $P(\underline{y}, t)$ by the relation

$$P(\Phi, t) = \int [P(\underline{y}, t)]_{y_0 = \Phi - y_1}\, dy_1\ . \qquad (5\text{-}10)$$

Substitution of (5-8) and (5-9) into (5-6) yields

$$a\frac{dy_0(t)}{dt} + aek\sum_{m=1}^{N(t)}\cos\theta'(\overline{y},\ t_m) +$$

$$\frac{dy_1(t)}{dt} + ek\cos\theta'(\overline{y}, t)\sum_{m=1}^{N(t)}\delta(t-t_m) = 0 \qquad (5\text{-}11)$$

where the dependence of $\theta'(t)$ on $\overline{y}(t)$ is emphasized. The above may be decomposed into the two equations

$$a\frac{dy_0(t)}{dt} + aek\sum_{m=1}^{N(t)} \cos\theta'(\bar{y}, t_m) = 0$$

$$\frac{dy_1(t)}{dt} + ek\cos\theta'(\bar{y}, t)\sum_{m=1}^{N(t)} \delta(t-t_m) = 0$$

where it is noted that the second equation is simply the derivative of the first. The above two equations may therefore be represented by the two first-order differential equations:

$$\frac{dy_0(t)}{dt} = y_1(t) \tag{5-12a}$$

$$\frac{dy_1(t)}{dt} = -ek\cos\theta'(\bar{y}, t)\sum_{m=1}^{N(t)} \delta(t-t_m) \quad . \tag{5-12b}$$

The above equations specify the dynamics of the vector process y(t) corresponding to (5-9). It is therefore possible to determine the equation for the joint density $P(\underline{y}, t)$ in (5-10) by using the two dimensional Smoluchowski in (5-3). The increments of the vector components given in (5-12) are

$$\Delta y_0 = y_1(t)\Delta t \tag{5-13a}$$

$$\Delta y_1 = -ek\int_t^{t+\Delta t} \cos\theta'(\bar{y}, \tau)\sum_{m=1}^{N(\tau)} \delta(\tau - \tau_m)\, d\tau$$

$$= -ek\sum_{m=1}^{N(\Delta t)} \cos\theta'(\bar{y}, t_m) \quad . \tag{5-13b}$$

These increments are needed to calculate the joint moments, $K_{mn}(y_0, y_1)$ given by Equation (5-4). The $K_{mn}(y_0, y_1)$ are calculated to be

$$K_{10}(\bar{y}) = y_1$$

$$K_{01}(\bar{y}) = -\frac{ekA}{2}\sin(y_0 + y_1)$$

$$K_{02}(\bar{y}) = (ek)^2(A)/2$$

$$K_{on}(\bar{y}) = (-ek)^n \cos^n \theta'(\bar{y}, t)\, n(t, \theta) \qquad n \geq 3$$

where again

$$n(t, \theta) = A + A\sin[\omega_s t + \theta_1(t)] \ .$$

All the other $K_{mn}(\bar{y})$ moments not listed above are zero. In this case the two-dimensional Smoluchowski series equation becomes

$$\frac{\partial p(\bar{y})}{\partial t} = -y_1 \frac{\partial p(\bar{y})}{\partial y_0} + \frac{\partial}{\partial y_1}\left[\frac{ekA}{2}\sin(y_0 + y_1)\, p(\bar{y})\right]$$

$$+ n\,\frac{(ek)^2 A}{4}\,\frac{\partial^2 p(\bar{y})}{\partial y_1^2}$$

$$+ \sum_{n \geq 3}^{\infty} \frac{(-1)^n}{n!}\,\frac{\partial^n}{\partial y_1^n}[K_{on}\, p(\bar{y})]. \qquad (5\text{-}14)$$

The above again represents an infinite order partial differential equation for the joint density $P(\underline{y}) \equiv P(\underline{y}, t)$. Some simplicity is afforded by considering only the steady state solution, but the resulting equation is still difficult to solve explicitly without digital computation.

For the case where the average intensity A is much greater than the loop bandwidth B_L (i.e., larg electron density) the approximate steady state solution to (5-14) can be found by limiting the number of terms involved. The corresponding steady state solution for $\Phi(t)$ from (5-10) is then approximately,

$$P(\Phi) \approx \frac{e^{\alpha \cos\Phi}}{2\pi I_0(\alpha)} \qquad A >> B_L \tag{5-15}$$

where $\alpha = A/2B_L$, but B_L now has the definition

$$B_L = \frac{eAk + 2a}{4} \quad . \tag{5-16}$$

That is, the loop bandwidth B_L is increased by the added zero in the loop filter of (5-1). The high density solution for $P(\Phi)$ is therefore identical to that of the first order loop case, with the adjustment in th B_L bandwidth. For higher-order loops an equivalent n-dimensional vector process must be defined and an n-dimensional Smoluchowski equation must be derived, increasing the complexity of the problem.

5.3 General Delay Tracking Loops

The objective of this section is to investigate the behavior of a phase tracking system when the input intensity modulation signal and loop feedback function are to a general periodic nature, but not necessarily sinusoidal. Let the signal electron rate of Equation (2-2) be represented by $n_s(t, \tau_1(t))$ and the feedback function by $y(t, \tau_2(t))$ where $\tau_1(t)$ and $\tau_2(t)$ are their respective time delays. The differential equation describing the loop operation for shot noise inputs, where θ_1 is again assumed constant, becomes

$$\frac{d\tau(t)}{dt} = -ek \sum_{m=1}^{N(t)} \delta(t-t_m)\, y(t, \tau_2(t)) \tag{5-17}$$

where $\tau(t) = \tau_1(t) - \tau_2(t)$ and N(t) is again a Poisson random variable with intensity $n_s(t, \tau_1(t))$. The incremental delay error is

$$\Delta\tau = -ek \sum_{m=1}^{N(\Delta t)} y(t_m, \tau_2(t_m)) \tag{5-18}$$

and the $K_n(\tau)$ moments of the Smoluchowski series equation are now

$$K_n(\tau) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E\left[-ek \sum_{m=1}^{N(\Delta t)} y(t_m, \tau_2(t_m)) \right] \tag{5-19}$$

where the expectation is conditioned on τ. Thus, Equation (5-19) becomes

$$K_n(\Phi) = (ek)^n \overline{y^n(t, \tau_2(t))\, n(t, \tau_1)} \tag{5-20}$$

where the over-bar represents time averaging inherent in the loop mixer function. Hence, $K_n(\Phi)$ is of the same form as in Equation (2-39), and would be identical to it if

$$n(t, \tau_1) = A + A\sin(\omega_0 t + \tau_1)$$

$$y(t, \tau_2(t)) = \partial n(t, \theta)/\partial\theta$$

$$= \cos(\omega_0 t + \tau_2(t)) \ .$$

The third-order truncated Smoluchowski series equation* for a general input function becomes

$$\frac{\partial p(\tau, t)}{\partial t} = ek \frac{\partial}{\partial \tau}\left[\overline{y(t, \tau_2(t))\, n(t, \tau_1)}\, p(\tau, t) \right] + \frac{(ek)^2 \partial^2}{2\partial \tau^2} \cdot$$

$$\left[\overline{y^2(t, \tau_2(t)\, n(t, \tau_1)}\, p(\tau, t) \right] + (ek)^3 \frac{\partial^3}{\partial \tau^3}\left[\overline{y^3(t, \tau_2(t))\, n(t, \tau_1)}\, p(\tau, t) \right] \tag{5-21}$$

*Here, attention is restricted again to only the first three terms of the infinite series equation, accepting the results as only an approximate solution.

The steady state version of Equation (5-21) occurs when the left-hand side is zero. Integrating the equation with respect to τ gives

$$C_0 = ek[R_{yn}(\tau)\, p(\tau)] + n\frac{(ek)^2}{2}\frac{d}{d\tau}[R_{y^2n}(\tau)] + (ek)^3\frac{d^2}{d\tau^2}[R_{y^3n}(\tau)\, p(\tau)] \qquad (5\text{-}22)$$

where

$$R_{yn}(\tau) = \overline{yn}$$

$$R_{y^2n}(\tau) = \overline{y^2n}$$

$$R_{y^3n}(\tau) = \overline{y^3n}$$

are correlation functions. Note that this equation corresponds to the previously considered equation (2-45) with the sinusoidal functions replaced by the general time averaged correlation functions given above.

5.4 An Example-Early Late Gate Tracking

In rader and pulse tracking systems a periodic pulse train is locked to a locally generated periodic signal through a feedback tracking system, similar to that in Figure 2. When the two signals are in time lock, the local signal tracks the time variations in the arrival times of the incoming periodic pulse train. In optical tracking systems the pulse train is generated by a pulsed laser whose intensity is detected by a photodetector at the receiver. The feedback signal in the tracking loop is designed such that when it is multiplied with the detected pulse train and integrated over some period the result is an error function that is odd with respect to the time delay. This local signal is often designed

to be a periodic train of positive-negative pulses as in Figure 13. The multiplication of the received pulse train by this particular local signal is equivalent to "gating in" the former signal by the latter signal. Hence, the receiver is often called an "early late gate" on "split-gate" tracker.

Since the output of a photodetector is a shot noise process with intensity $n(t, \tau_1)$, the analysis problem is an example of the application of the general tracking theory of the previous section. By referring to Figure 13 it is easily seen that

$$y^3(t, \tau_2(t)) = y(t, \tau_2(t))$$

$$y^2(t, \tau_2(t)) = \frac{1}{2A}[n(t, \tau_2(t))]$$

and therefore

$$R_{yn}(\tau) \equiv \overline{y(t, \tau_2(t))\, n(t, \tau_1)} = \overline{y^3(t, \tau_2(t))\, n(t, \tau_1)}$$

$$= R_{y^3 n}(\tau)$$

$$= \frac{1}{T}\int_0^T y(t, \tau_2(t))\, n(t, \tau_1)\, dt \qquad (5\text{-}23)$$

$$R_{nn} \equiv \overline{y^2(t, \tau_2(t))\, n(t, \tau_1)}$$

$$= \frac{1}{2AT}\int_0^T n(t, \tau_2(t))\, n(t, \tau_1)\, dt + (\text{a constant}) . \qquad (5\text{-}24)$$

Equation (5-22) now becomes

$$C_0 = ek\, R_{yn}(\tau)\, p(\tau) + \frac{(ek)^2}{2}\frac{d}{d\tau}[R_{nn}(\tau)\, p(\tau)] + \frac{(ek)^3}{6}\frac{d^2}{d\tau^2}[R_{yn}(\tau)\, p(\tau)] \qquad (5\text{-}25)$$

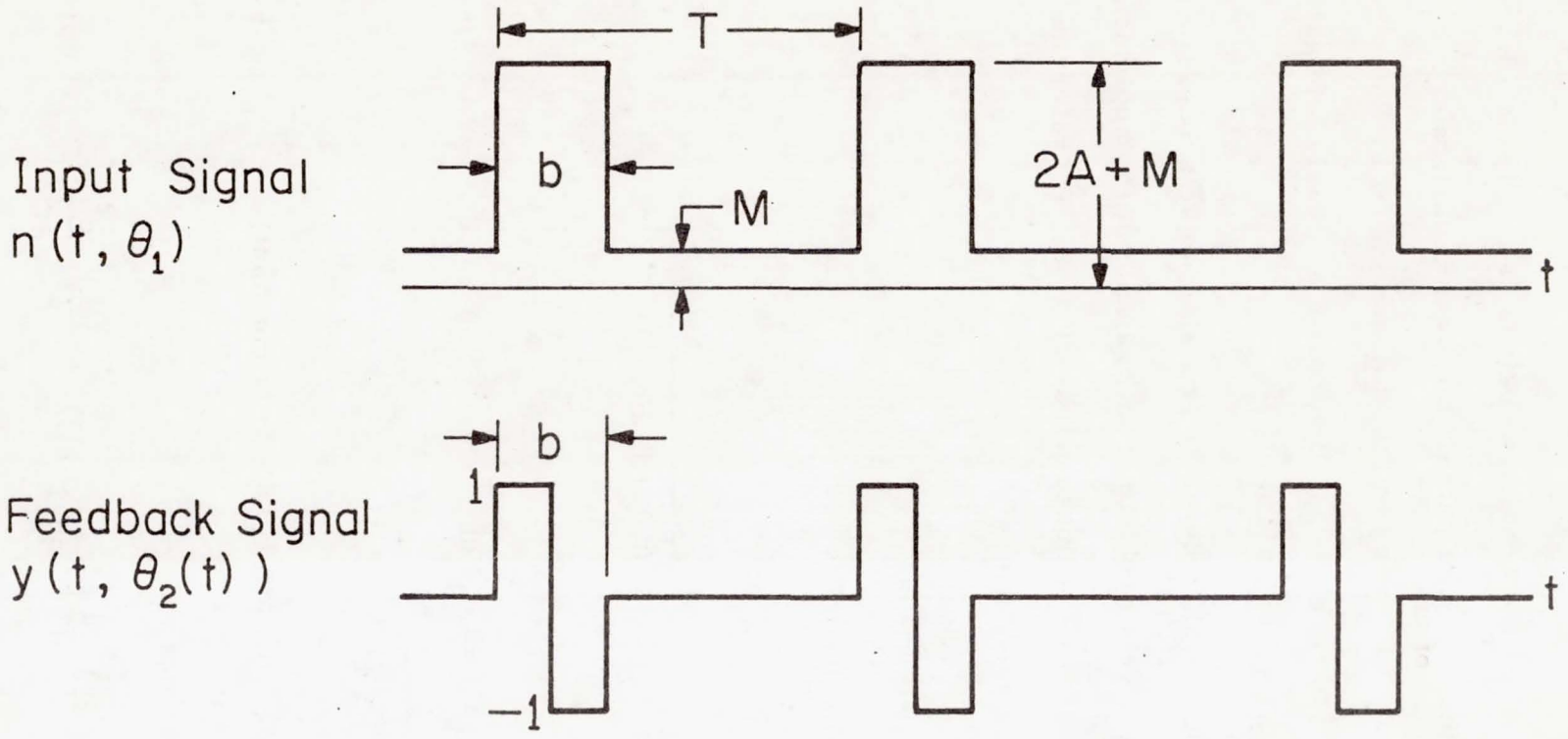


Figure 13.

Therefore, for the given input and feedback functions in Figure 13, Equation (5-25) could be solved as a second-order differential equation for the delay error density of an "Early Late Gate" tracking system. The solution would represent an approximate solution to the infinite series Smoluchowski equation. A computer solution similar to that used in Chapter 3 would be applicable to the solution of Equation (5-25).

REFERENCES

1. Viterbi, A. J., Principles of Coherent Communication, McGraw-Hill, Inc. (1966).

2. Davenport, Jr., W. B., and W. L. Root, Random Signals and Noise, McGraw-Hill, Inc. (1958).

3. Middleton, D., Introduction to Statistical Communication Theory, McGraw-Hill, Inc. (1960).

4. Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill, Inc. (1965).

5. O'Niell, E. and Karp, S. and Gagliardi, R., "Communication Theory for the Free Space Optical Channel", Proc. of the IEEE, October 1970.

6. Karp, S., "A Statistical Model for Radiation with Applications to Optical Communications", Ph. D. Dissertation, University of Southern California (Jaunary 1967).

7. Anderson, L. K., and B. J. McMurty, "High Speed Photodetectors", Proceedings of the IEEE, Volume 54, Number 10 (October 1966).

8. Reiffen, B., and H. Sherman, "An Optimum Demodulator for Poisson Processes; Photo Source Detectors", Proceedings of the IEEE, (October 1963).

9. Bar-David, Israel, "Communication Under Poissonian Regime", Israel Ministry of Defense, Scientific Department, Report No. 40/07-526 (January 1968).

10. Parzen, E., Stochastic Processes, Holden Day, Inc. (1962).

11. Lindsey, William C., "Nonlinear Analysis and Synthesis of Generalized Tracking Systems", USCEE Technical Report 317, (December 1968).

12. Pratt, William K., Laser Communication Systems, John Wiley & Sons, Inc., (1969).

13. Coddington, Earl A., and Norman Levinson, Theory of Ordinary Differential Equations, McGraw-Hill Book Company, Inc. (1955).

14. Ince, E. L., Ordinary Differential Equations, Dover Publications, Inc., (1956).

15. Boyce, William E., and Richard C. DiPrima, Elementary Differential Equations and Boundary Value Problems, John Wiley and Sons, Inc., (1965).

16. Kolmogorov, A.N., "On Analytical Methods in Probability Theory", Math. Ann., Vol. 104, pp. 415-458 (in German), (1931).

17. Stratonovich, R.L., Topics in the Theory of Random Noise, Vol. 1, New York: Gordon and Breach, (1963).

January 1971

USCEE Report 401

Interim Technical Report

Counting Statistics for Extended Optical Photodetectors

R. Gagliardi

U. Farrukh

Department of Electrical Engineering

University of Southern California

Los Angeles, California 90007

This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-104. This grant was part of the research program at NASA's Goddard Space Flight Center, Greenbelt, Maryland.

Page intentionally left blank

ABSTRACT

Recent attention has been devoted to the derivation of the probability density of the number of photoelectrons occurring (count density) at the output of a photodetecting surface when receiving an optical field. Previous analysis has been basically confined to point detectors, in which the spatial extension of the detector surface over the normal optical field has been ignored. In this report count density analysis is generalized to account for the extended detector. The accepted mathematical detector model is developed to cover both time and spatial counting, and the integral equations necessary for exact analysis are developed. Approximate solutions to these equations are presented for some cases of practical interest. Knowledge of counting statistics is of utmost importance in the optimal design of an optical communication or tracking system.

1.0 Introduction

In optical systems, knowledge of the statistics of a photodetecting receiver is necessary for the application of optimum detection and estimation procedures. The statistic of most importance in an optical communication system is the probability of the number of photoelectrons produced at the photodetector output when receiving a statistical field. Although this detection operation is basically quantum mechanical in nature, the density of the electron occurrences can be theoretically derived by using a semi-classical approach to the detector operation. This method can be used for developing probabilistic models while avoiding the basic physics underlying the receiver. In past work, the problem of determining photoelectron probabilities has been almost exclusively confined to photodetectors in which the spatial extension of the detector has been ignored, and only temporal effects have been developed. In this report the above approaches are extended to include the spatial effects of the detector during the photodetection operation.

2.0 The Optical Photodetector

An optical detector is a photosensitive surface that responds to incident optical radiation by releasing electrons. These electrons are captured by an anode plate, producing an electron current. The release of the electrons is strongly influenced by the incident optical field, but basically behaves as a random phenomena. The resulting electron current therefore evolves as a random process, and is mathematically modeled as shot noise whose statistical behavior is directly related to the number of electrons produced at the anode. The number of electrons produced is called the electron "count" and its associated statistical properties are referred to as "count statistics". Of particular significance is the probability density of the number of electrons produced during a given time interval from the entire detector surface, when receiving a given optical field.

The accepted mathematical model[1] of a photodetector is derived using a semi-classical approach, which treats the electromagnetic field classically, but prescribes a probabilistic solution to account for its interaction with the atomic structure of the detector surface. Although a complete description of the emission and absorbtion of light by an atom influenced by a radiation field is well beyond the scope of this report, an outline of the approach as it is related to count statistics is presented below.

The semi-classical derivation begins with the Hamiltonian equations for a charged particle in an electromagnetic field. It is then assumed that the combined system, atom plus radiation, begins in some initial state, and a set of coupling equations are derived for the transition probabilities, from which one can determine the probability rate of finding the combined system in a

given final state. Summing over all final states, and making some simplifying assumptions, one ends with Fermi's rule for the probability per second for a state transition over a differential area $\Delta\underline{r}$ located at point $\underline{r}$ on the detector surface.

$$\frac{dP(t)}{dt} = \beta I(\underline{r},t)\ \Delta\underline{r} \tag{1}$$

Here, β is a proportionali v constant and $I(\underline{r},t)$ is the received normal electromagnetic power at time t and point $\underline{r}$ on the detector surface. The primary consequence of the Fermi rule is that it implies that in a short time interval Δt, the probability of ejecting an electron from an atom at the elemental surface area $\Delta\underline{r}$ is proportional to the incident radiation energy over Δr and Δt. That is,

$$\left[\begin{array}{l}\text{Probability that an electron} \\ \text{is released from area } (\underline{r} + \Delta\underline{r}) \\ \text{during time interval } (t,\ t + \Delta t).\end{array}\right] \cong \beta I(\underline{r},t)\Delta\underline{r}\Delta t \tag{2}$$

for sufficiently small $\Delta\underline{r}$ and Δt. In addition, (2) implies that the probability of more than one electron being released must go to zero as $(\Delta r \Delta t)^2$, which means

$$\left[\begin{array}{l}\text{Probability no electron } \textit{is} \\ \text{released from area } (\underline{r} + \Delta\underline{r}) \\ \text{during time interval } (t,\ t + \Delta t).\end{array}\right] \cong 1 - \beta I(\underline{r},t)\Delta r\Delta t \tag{3}$$

as $\Delta r \Delta t \to 0$. Note that (2) states that the release of an electron from any elemental area at $\underline{r}$ at any time t depends only upon the radiation energy at that time and point, which implies that the release of electrons from disjoint differential areas on the surface, and from disjoint time intervals, can be treated as independent events. This assumption, along with Equations (1) and (3), describe the mathematical model of the photodetecting surface, and will be of primary importance in the subsequent derivation of the total electron count.

3.0 Derivation of Count Densities

The probability density of the number of electrons produced during a time interval $(t, t + \tau)$ over a total surface area A can be derived using the photodetector model of the previous section. To facilitate this derivation, we introduce the notion of a time-space domain. This domain contains vectors whose components correspond to time and spatial coordinates associated with the detector surface. For simplicity, we write these vectors as $\underline{v} = (t,\underline{r})$ where t is the scalar time component and $\underline{r}$ represents the two-dimensional spatial coordinates of the detector surface. We define the volume V in this domain to be composed of all vectors $\underline{v}' = (t',\underline{r}')$ such that $t \leq t' \leq t + \tau$ and $\underline{r}' \in A$, where A is the spatial area encompassed by the detector surface area and τ is the counting time. This allows us to denote the normal electromagnetic field intensity (power) at point r' on the detector surface at time t' by $I(t',\underline{r}') \equiv I(v')$. This means the volume V is basically the set of all points in the time-space domain over which we observe the radiation field with a given detector in a given time interval.

Now consider the partition of the volume V into disjoint cells $\Delta V = \Delta\underline{r}\Delta t$. We shall assume $\Delta\underline{r}$ and Δt are smaller than the spatial and time variations in $I(t,\underline{r})$ to insure that within each ΔV, $I(\underline{v})$ is approximately constant. (This is always possible with continuous fields.) Let Δv be the volume of the cells ΔV, and let η be the total number of cells in V after partitioning. The ensemble of η disjoint cells ΔV can now be ordered to form the sequence $\{\Delta V_1, \Delta V_2, \ldots \Delta V_\eta\}$ where each ΔV_i is centered at some point v_i in V. Note that each ΔV_i can be interpreted as an observation cell corresponding to an elemental surface area and elemental time interval over

which we observe the radiation field. In this notation, the model of the detector in the previous section has the property:

$$\left[\begin{array}{l}\text{Probability of an electron}\\ \text{emitted from } \Delta V_i \text{ at } v_i.\end{array}\right] = \beta I(v_i)\Delta v \qquad (4)$$

We now consider the probability of the detector releasing k total electrons from the total surface area A during the total time interval $(t,t+\tau)$. This is equivalent to the compound probability that k electrons are emitted from the totality of all cells $\{\Delta V_i\}$ spanning V. This can be written as

$$P(k,V) = \frac{1}{k!} \sum_{\substack{\text{all}\\ \text{orderings}}} \left[\begin{array}{l}\text{Probability of}\\ \text{one electron}\\ \text{from } k \text{ different}\\ \text{ordered cells.}\end{array}\right] \left[\begin{array}{l}\text{Probability of}\\ \text{no electrons in}\\ \text{the } \eta\text{-}k \text{ remain-}\\ \text{ing ordered cells.}\end{array}\right]$$

$$\cong \frac{\beta^k}{k!} \sum_{\substack{\text{all}\\ \text{orderings}}} I(v_{i_1})\ldots I(v_{i_k})(\Delta v)^k \prod_{q=k+1}^{\eta} \left[1-\beta I(v_{i_q})\Delta V\right] \qquad (5)$$

where the summation must consider all possible orderings of the η cells without repeats, and the division by $k!$ is necessary since particular arrangements involving the same k cells need only be considered once. Note that (5) has used (3), (4), and the assumption of independent electron emissions from disjoint cells.

We are interested in the limit of (5) as $\Delta v \to 0$, $\eta \to \infty$, so that the approximation indicated becomes a true equality. Since the limits of sums and products is equal to the sum and products of the limits, we can investigate the limit of the individual terms in (5) and recombine. We first show that the product term has the same limit for all orderings. This can be seen by considering the limit of the logarithm of the product. The log is

$$\log \prod_{q=k+1}^{\eta} \left[1 - \beta I(v_{i_q})\Delta v\right] = \sum_{q=k+1}^{\eta} \log \left[1 - \beta I(v_{i_q})\Delta v\right] \tag{6}$$

Adding and subtracting the terms omitted in each ordering allows us to always write (6) as

$$\sum_{q=1}^{\eta} \log \left[1 - \beta I(v_{i_q})\Delta v\right] - \sum_{q=1}^{k} \log \left[1 - \beta I(v_{i_q})\Delta v\right] \tag{7}$$

Now, in the limit as $\Delta v \to 0$, $\eta \to \infty$ and

$$\lim_{\Delta v \to 0} \log [1 - \beta I(v_i)\Delta v] \to -\beta I(v_i)\Delta v \tag{8}$$

The first summation in (7) therefore has the limit

$$\lim_{\substack{\Delta v \to 0 \\ \eta \to \infty}} \sum_{q=1}^{\eta} \log [1 - \beta I(v_q)\Delta v] \to \lim_{\substack{\Delta v \to 0 \\ \eta \to \infty}} \beta \sum_{q=1}^{\eta} [-I(v_q)\Delta v]$$

$$\to \beta \int_V I(v)dv \tag{9}$$

while the second summation in (7) involves only a finite number of terms, has the limit

$$\lim_{\substack{\Delta v \to 0 \\ \eta \to \infty}} \beta \sum_{q=1}^{k} \left[-I(v_{i_q}) \Delta v \right] \to 0 \tag{10}$$

for each possible ordering. Hence, in (5)

$$\lim_{\Delta v \to 0} \prod_{q=k+1}^{\eta} \left[1 - \beta I(v_{i_q}) \Delta v \right] \to \exp \left[-\beta \int_V I(\underline{v}) d\underline{r} \right] \tag{11}$$

Now consider the summation term in (5):

$$\lim_{\Delta v \to 0} \beta^k \sum_{\substack{\text{all} \\ \text{orderings}}} \left[I(v_{i_1}) \ldots I(v_{i_k}) (\Delta v)^k \right] \tag{12}$$

Since each ordering above requires $i_1 \neq i_2 \neq \ldots \neq i_k$, the summation can be written as

$$\beta^k \sum_{i_k=1}^{\eta} \ldots \sum_{i_2=1}^{\eta} \sum_{i_1=1}^{\eta} \left[\bullet, \bullet, \bullet \right] - \left[\begin{array}{l} \text{Sum over all} \\ \text{orderings in} \\ \text{which at least} \\ \text{two } i_\alpha \text{ are} \\ \text{equal.} \end{array} \right] \tag{13}$$

The second term will involve η^{k-1} terms of order $(\Delta v)^k$. Since η behaves as $1/\Delta v$, the limit of the second term will be zero as $\Delta v \to 0$. The first term in (13), however, has the limit

$$\lim_{\Delta v \to 0} \left(\beta \sum_{i=1}^{\eta} I(v_i) \Delta v \right)^k \to \left(\beta \int_V I(v) dv \right)^k \tag{14}$$

Therefore, using (9), (10), and (14) in (5), we derive the desired probability of k emissions over V as

$$P(k \quad V) = \frac{(m_v)^k}{k!} \quad \exp(-m_v) \tag{15}$$

where

$$m_v = \beta \int_V I(v) d\underline{v} = \int_t^{t+\tau} \int_A \boldsymbol{n}(t,\underline{r}) dt d\underline{r} \tag{16}$$

and

$$\boldsymbol{n}(t,\underline{r}) \equiv \beta I(t,\underline{r}) = \beta \left| f(t,\underline{r}) \right|^2$$

The probability that exactly k electrons will be emitted is therefore related to the integral of the squared field envelope over the desired observation volume. The parameter k can represent any non-negative integer, and therefore (15) represents a probability over all integers $(0,\infty)$. Note that the normalized intensity $\boldsymbol{n}(t,\underline{r})$ effects the count probability only through the functional m_v. We have explicitly indicated the dependence of the probability on the counting volume V, which in turn depends upon the location and size of the detector area A, the counting interval τ, and in particular on the specific time parameter t. Thus, the probability in (15) is in general

non-stationary in time. This dependence upon time and spatial area A will be an important aspect in subsequent analysis.

If the optical field $f(t,\underline{r})$ is constant over A at any time t (i.e., $f(t,\underline{r}) = r_o f(t)$ for all $\underline{r} \in A$), then the spatial effects can be removed, and (16) becomes

$$m_v = \beta \int_A \int_t^{t+\tau} \left| r_o f(t) \right|^2 dt d\underline{r} = \beta a r_o^2 \int_t^{t+} \left| f(t) \right|^2 dt \qquad (17)$$

where a is the integrated area over the detector surface and r_o^2 is the field power per unit area. The detector is said to be a point-detector, since it basically collects power at a particular point in space, normalized by area power ar_o^2. This latter effect can be incorporated into the coefficient β, which can now be redefined as

$$\alpha \equiv a r_o^2 \beta \qquad (18)$$

Thus, when dealing with point detectors in (15), the previously defined time-space intensity $n(t,\underline{r}) = \beta |f(t,\underline{r})|^2$ can be replaced by the time intensity function $n(t) = \alpha |f(t,\underline{r})|^2$ in (16).

4.0 The Karhunen-Loève Expansion of the Field

The key to further investigation of (15), using (16), depends upon the ability to expand the radiation field $f(t,\underline{r})$ into orthonormal Fourier series over the observation volume V:

$$f(\underline{v}) = \sum_{i=1}^{\infty} f_i \phi_i(\underline{v}) \tag{19a}$$

where

$$f_i = \int_V f(\underline{v})\phi_i(\underline{v})d\underline{v} = \int_A \int_t^{t+\tau} f(t,\underline{r})\phi_i(t,\underline{r})dtd\underline{r} \tag{19b}$$

are the complex Fourier coefficients and $\{\phi_i(\underline{v})\}$ represents a complete set of complex orthonormal bases functions over $\underline{v} \in V$. That is,

$$\int_V \phi_i(\underline{v})\phi_j^*(\underline{v})d\underline{v} = \int_A \int_t^{t+\tau} \phi_i(t,\underline{r})\phi_j(t,\underline{r})dtd\underline{r} = \delta_{ij} \tag{20}$$

If $f(t,\underline{r})$ is squared integrable over V, then the equality in (19a) is in the squared integrable sense, and the convergence requires only a bounded energy constraint on the radiation field over V. When $f(\underline{v})$ is a stochastic fiel , the coefficients $\{f_i\}$ become complex random variables, and the representation in (19a) is in a mean square sense. In the latter case, if the orthonormal set $\{\phi_i(\underline{v})\}$ is selected so that

$$\int_V K_f(\underline{v}_1,\underline{v}_2)\phi_i(\underline{v}_1)d\underline{v}_1 = \gamma_i\phi_i(\underline{v}_2) \tag{21}$$

where the Kernel

$$K_f(\underline{v}_1,\underline{v}_2) \equiv E\left[f(\underline{v}_1)f^*(\underline{v}_2)\right]$$

$$= E\left[f(t_1,\underline{r}_1)f^*(t_2,\underline{r}_2)\right] \tag{22}$$

Then the random coefficients $\{f_i\}$ are uncorrelated, and the expansion in (19a) is called a Karhunen-Loeve (K-L) expansion. The function $K_f(\underline{v}_1,\underline{v}_2)$ is the covariance function of the radiation field, and is obtained by averaging in (22) over the statistics of the field. The convergence in (22) then requires only the squared-integrability and continuity of the covariance function $K_f(\underline{v}_1,v_2)$ over V. The orthonormal set $\{\phi_i(\underline{v})\}$ that are solutions to the integral equation in (21) are called eigenfunctions of the integral operator with kernal $K_f(\underline{v}_1,\underline{v}_2)$, and the γ_i are the associated eigenvalues. These eigenvalues of the K-L expansion are particularly important to our interests since

$$E\left|f_i\right|^2 = E\int_V f(\underline{v}_1)\phi_i(\underline{v}_1)d\underline{v}_1 \int_V f^*(\underline{v}_2)\phi_i(\underline{v}_2)d\underline{v}_2$$

$$= \int_V\int_V E\left[f(\underline{v}_1)f^*(\underline{v}_2)\right]\phi_i(\underline{v}_1)\phi_i(\underline{v}_2)d\underline{v}d\underline{v}_2$$

$$= \int_V \gamma_i\phi_i(\underline{v}_1)\phi_i(\underline{v}_1)d\underline{v}_1 \tag{23}$$

$$= \gamma_i$$

where we have used (19b), and the fact that $\{\phi_i(\underline{v})\}$ are orthonormal over V. Thus, the eigenvalue γ_i is the mean square value of the random variable f_i.

With the K-L expansion in (19) we can now substitute into (16) and derive

$$\begin{aligned} m_V &= \int_V \beta \left| f(\underline{v}) \right|^2 d\underline{v} \\ &= \int_V \beta \left[\sum_{i=1}^{\infty} f_i \phi_i(\underline{v}) \right] \left[\sum_{j=1}^{\infty} f_j^* \phi_j^*(\underline{v}) \, d\underline{v} \right] \\ &= \beta \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_i f_j^* \int_V \phi_i(\underline{v}) \phi_j(\underline{v}) d\underline{v} \end{aligned} \tag{24}$$

where the interchange of summation and integration is guaranteed by the convergence of the series. The orthonormality of the basic functions then yield

$$m_V = \beta \sum_{i=1}^{\infty} \left| f_i \right|^2 \tag{25}$$

Thus, we have expressed the counting level as the sum of the magnitude squared of the coefficients. Note that the particular observation volume V over which we collect radiation, and which depends upon the detector area A and the counting time interval $(t,t+\tau)$, is implicit in the computation of the coefficients on the right.

The expansion of m_V in (25) as a sum of random variables affords us a convenient systematic approach for determining the count probabilities in

(15). We can either attempt to compute the probability density of the sum in (25), and evaluate (15), or alternatively, compute the characteristic function $\psi_m(\omega)$ of the sum in (25), and evaluate $p(k,V)$ by [2]

$$p(k,V) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_m(e^{j\omega} - 1)\, e^{-j\omega k}\, d\omega \tag{26}$$

The computation of the characteristic function $\psi_m(\omega)$ however requires joint statistics over all the $\{f_i\}$. (The K-L expansion produces uncorrelated coefficients, but the joint density is still needed.) By considering only Gaussian fields, this latter obstacle is avoided. This is due to the fact that uncorrelated Gaussian random variables are independent. A Gaussian field is a stochastic field $f(t,\underline{r})$ and is a Gaussian random variable at every t and $\underline{r}$. For Gaussian fields, the K-L expansion has the added feature of producing independent Gaussian random coefficients f_i in (19). It then follows that

$$\psi_m(\omega) = \prod_{i=1}^{\infty} \left[\frac{1}{1-\beta\gamma_i\omega}\right] \exp\left[\frac{\beta|s_i|^2\omega}{1-\beta\gamma_i\omega}\right] \tag{27}$$

where s_i is the mean (signal) value of f_i and γ_i is the eigenvalue in (21). Note the above characteristic function appears as a product of individual characteristic functions. This means it can be enterpreted as the characteristic function of a sum of random variables $\{k_i\}$, where each k_i has the probability density

$$p(k_i) = \frac{(\beta\gamma_i)^{k_i}}{(1+\beta\gamma_i)^{1+k_i}} \exp\left[\frac{-\beta|s_i|^2}{1+\beta\gamma_i}\right] L_{K_i}\left[-\frac{|s_i|^2}{\gamma_i(1+\beta\gamma_i)}\right] \tag{28}$$

and $L_K(\cdot)$ is the k^{th} order Laguerre density in argument $(\cdot)$. Thus, for stochastic Gaussian fields, the count statistic $\underline{K}$ can be interpreted as the sum

$$\underline{K} = \sum_{i=1}^{\infty} K_i \tag{29}$$

of independent random counts k_i, where each k_i has a Laguerre count probability given by (28). Each k_i can be considered the random count associated with a time-space mode of the Gaussian field. The modes, however, have specific meaning, since they must be associated with the particular orthonormal set of the K-L expansion. Each such mode contributes an independent Laguerre count variable to the total count. Note that if $s_i = 0$ for some i, implying no deterministic component in that particular mode, (28) becomes

$$p_{k_i}(k) = \frac{(\beta\gamma_i)^k}{(1+\beta\gamma_i)^{1+k}} \tag{29}$$

which is the Bose-Einstein probability. Hence, any mode of the Gaussian field that does not contain a signal component (zero mean value) contributes a Bose-Einstein count variable to $\underline{K}$ in (29). We emphasize that the Laguerre densities in (28), and the Bose-Einstein densities in (29), require knowledge of the eigenvalues $\{\gamma_i\}$, which in turn require solution of the multi-dimensional integral

equation (21) associated with the K-L expansion. For certain special types of stochastic Gaussian fields, the solution is somewhat simplified, as discussed below.

5.0 Stationary, Homogenious, and Coherence-Seperable Stochastic Fields

A stochastic field is said to be temporally stationary if the time dependence in coherence function $K_f(\underline{r}_1,\underline{r}_2;t_1t_2)$ depends only upon the time difference t_1-t_2. The field is spatially homogenious if the spatial dependence in the coherence function depends only upon the spatial distance $(\underline{r}_1-\underline{r}_2)$. A field is completely homogenious if it is both temporally stationary and spatially homogenious.

A stochastic field is said to be a coherence-separable field if its coherence function factors a

$$K_f(\underline{r}_1,\underline{r}_2;t_1t_2) = K_s(\underline{r}_1,\underline{r}_2)\ K_t(t_1,t_2) \tag{30}$$

The factor $K_s(\underline{r}_1,\underline{r}_2)$ is then called the spatial (mutual) coherence function, while $K_t(t_1,t_2)$ is called the covariance function of the field. A field that is coherence-separable and temporally stationary [i.e., $K_t(t_1,t_2) = K_t(t_1,t_2)$] is said to be a spectrally pure field.

An optical field is completely space coherent over an area A if $K_s(\underline{r}_1,\underline{r}_2) = 1$ for all $\underline{r}_1,\underline{r}_2$ in A. It is completely space incoherent if $K_s(\underline{r}_1,\underline{r}_2) = 0$, $\underline{r}_1 \neq \underline{r}_2$. Otherwise, it is partially space coherent. For coherence-separable fields, the K-L expansion has the property that the left side of the eigenfunction equation in (21) becomes

$$\int_A \int_t^{t+\tau} K_s(\underline{r}_1,\underline{r}_2)K_t(t_1,t_2)\phi_i(t_1,\underline{r}_1)dt_1 d\underline{r}_1$$

$$= \int_A K_s(\underline{r}_1,\underline{r}_2) \int_t^{t+\tau} K_t(t_1,t_2)\phi_i(t_1,\underline{r}_1)dt_1 d\underline{r}_1 \qquad (31)$$

The form of this shows that an eigenfunction solution to (21) will occur as

$$\phi_i\ (t,\underline{r}) = g_i(t)h_j(\underline{r}) \qquad (32)$$

$$\gamma_i = \gamma_{it}\gamma_{js} \qquad (33)$$

where the above terms satisfy

$$\int_t^{t+\tau} K_t(t_1,t_2)g_i(t_1)dt_1 = \gamma_{it}g_i(t_2) \qquad (34)$$

$$\int_A K_s(\underline{r}_1,\underline{r}_2)h_j(\underline{r}_1)d\underline{r}_1 = \gamma_{js}h_j(\underline{r}_2) \qquad (35)$$

That is, for coherence-separable stochastic fields the eigenfunctions and eigenvalues will factor into a product of time and spatial components. Thus, the eigenfunctions over V in (21) can be determined by separately determining the family of eigenfunctions for both the time and space coherence kernals in (34) and (35). Since the product of any two such time and space eigenfunctions will satisfy (31), the set of all eigenfunctions $\{\phi_i(\underline{v})\}$ in (19) must involve all

possible pairwise products of the $\{g_i(t)\}$ and $\{h_j(\underline{r})\}$. The K-L expansion then takes the form

$$f(t,\underline{r}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{ij} g_i(t) h_j(\underline{r}) \tag{36a}$$

where

$$f_{ij} = \int_A \int_t^{t+\tau} f(t,\underline{r}) g_i(t) h_j(\underline{r}) dt d\underline{r} \tag{36b}$$

The functions $\{g_i(t)\}$ define the temporal modes of the received field, while the functions $\{h_j(\underline{r})\}$ designate its spatial modes. The corresponding $\{\gamma_{it}\}$ and $\{\gamma_{js}\}$ eigenvalues are the mean square power in each mode. If the field is completely spatially coherent over A, such that $K_s(\underline{r}_1,\underline{r}_2) = r_o^2$ over all $\underline{r}_1,\underline{r}_2$ in A, then (35) is satisfied with the single eigenfunction $h(\underline{r}) = 1$ and eigenvalue $\gamma_s = r_o^2 a$. Thus, the field will have only one spatial mode in A. Substitution into (36) also shows that $f(t,\underline{r}) = r_o^2 a f(t)$, as was assumed in the definition of the point detector in Section 3. The point detector assumption therefore is valid for stochastic fields whenever the field is completely spatial coherent over the detector area, or equivalently, when only one spatial mode exists.

6.0 Solutions of the Spatial Integral Equation

Computation of the count statistics require the solution to the integral equations in (34) and (35). Both equations are basically identical in form, although the former involves the scalar parameter t, while the latter involves the space vector $\underline{r}$. Solutions to the time equation (34) have been considered in the literature.[2] In this section we derive some approximate solutions for the spatial equation in (36).

A. Rectangular Detector

Consider a rectangular photodetector of length a, width b, and area $A = ab$, receiving a coherence separable field from a source at a distance R. Using the Fresnel-Kirckoff approximation, the spatial coherence function over the detector surface A is given by(1)

$$K_s(\underline{r}_1,\underline{r}_2) = [A\beta(o)]^{-1}\,\beta(\underline{r}_2-\underline{r}_1)\,\exp\left[j\left(\frac{\pi}{R\lambda}\right)\left(\underline{r}_1^2-\underline{r}_2^2\right)\right] \tag{37}$$

where λ is the wavelength, A is the area of the detector, and $\beta(\underline{r})$ has a Fourier Transform called the source radiance $B(u)$:

$$\beta(\underline{r}) = \int B(\underline{u})\exp[j2\pi(\underline{u}\cdot\underline{r})/\lambda R]d\underline{u} \tag{38}$$

Here the integration is over two-dimensional space of $\underline{u}$ vectors. By writing

---

(1) Helstron, C. W., J. Opt. Soc. Am., Vol. 59, March 1969
"Detection of Objects Through Turbulent Medium", p. 333

the eigenfunctions in (35) as

$$h_j(\underline{r}) = \hat{h}_j(\underline{r}) \exp\left[-j\pi\underline{r}^2/\lambda R\right] \tag{39}$$

we can rewrite the spatial integral equation in (36) as

$$\gamma_{js}\hat{h}(\underline{r}_2) = [A\beta(o)]^{-1} \int_A \beta(\underline{r}_2-\underline{r}_1)\, \hat{h}(\underline{r}_1)d\underline{r}_1 \tag{40}$$

Now, if the source subtends from the detector a solid angle much less than λ^2/A, then $\beta(\underline{r}) \approx \beta(o)$ over the detector area, and the detected field possesses complete coherency. Thus, the point detector model of the previous section is valid, and the source can be considered a point source. When the source is so large that its solid angle, when viewed from the detector, spans many multiples of λ^2/A, its coherence function over the detector surface area is generally smaller in area than the detector itself. For the rectangular detector, the eigenfunctions in (40) can be approximated by two-dimensional spatial plane waves

$$\hat{h}_j(\underline{r}) = \exp\left[j2\pi qx/a + j2\pi\ell y/b\right] \tag{41}$$

where x and y are components of $\underline{r}$, and q, ℓ are integers. That is, the eigenfunctions can be taken as spatial harmonics, with spatial frequencies that are multiples of the period a and b. If we substitute with (38) and (41), the right side of (40) becomes

$$[A\beta(o)]^{-1} \int_A \int_{u_2} B(\underline{u}) \exp\left[j2\pi \underline{u} \cdot (\underline{r}_2 - \underline{r}_1)/\lambda R\right] \exp\left[j2\pi q x_1/a + j2\pi \ell y_1/b\right] d\underline{r}_1 d\underline{u}$$

$$= [A\beta(o)]^{-1} \int_{u_2} B(\underline{\mu}) \exp\left[j \frac{2\pi(\cdot \underline{r}_2}{\lambda R}\right] d\underline{u}$$

$$\int_A \exp\left[j2\pi x\left(\frac{q}{a} - \frac{u_1}{\lambda R}\right) + j2\pi y\left(\frac{\ell}{b} - \frac{u_2}{\lambda R}\right)\right] dxdy \qquad (42)$$

where u_1 and u_2 are the components $\underline{u}$. The right hand integral integrates to approximately

$$\delta\left(\frac{q}{a} - \frac{u_1}{\lambda R}\right) \delta\left(\frac{\ell}{b} - \frac{u_2}{\lambda R}\right) \qquad (43)$$

so that (42) becomes, approximately,

$$\approx\left[\frac{\lambda^2 R^2}{A\beta(o)}\right] B\left(\frac{q\lambda R}{a}, \frac{\ell\lambda R}{b}\right) \exp\left[j \frac{2\pi q}{a} x_2 + j \frac{2\pi \ell}{b} y_2\right] \qquad (44)$$

where $B(u_1,u_2) = B(\underline{u})$. Thus, substituting (44) and (41) into (40), we have

$$\gamma(q,\ell) = \left[\frac{\lambda^2 R^2}{A\beta_o}\right] B\left(\frac{q\lambda R}{a}, \frac{\ell\lambda R}{b}\right) \qquad (45)$$

Therefore, the spatial eigenvalue associated with each of the two-dimensional harmonic spatial frequency $(q/a, \ell/b)$ is given by the evaluation of the

source radiance function $B(u_1,u_2)$ at the point $(u_1 = qa_o, u_2 = \ell b_o)$ where $a_o = \lambda R/a$, $b_o = \lambda R/b$. Thus, a significant spatial frequency harmonic, or spatial mode, will exist for all (q,ℓ) combinations such that (45) is non-negligible. Note the above is equivalent to effectively sampling $B(\underline{u})$ at points separated by $a_o = \lambda R/a$ in the u_1 dimension and by $b_o = \lambda R/b$ in the u_2 dimension. This means that if s_o is the two-dimensional spatial area over which the transform of $\beta(\underline{r})$ is non-negligible, then the number of significant spatial modes (harmonics) can be determined by partitioning s_o into grids, or disjoint squares of width a_o and b_o and determining the number of such squares needed to cover s_o. This means

$$[\text{number of spatial modes}] \simeq \frac{\text{area } s_o}{a_o b_o} + 1 \tag{46}$$

where the one factor is needed to include the case where $s_o << a_o b_o$. Note that the area $a_o b_o$ when projected to the source at distance R subtends at the detector a solid angle $a_o b_o/R^2 = (\lambda R/a)(\lambda R/b)/R^2 = \lambda^2/ab = \lambda^2/A$, which is the diffraction limited field of view of the detector. Therefore, the number of significant spatial modes over a rectangular detector surface can alternatively be viewed as the number of solid angles λ^2/A needed to cover the solid angle subtended by the radiance function $B(\underline{u})$ located at the source. Thus, we can also write

$$\text{number of spatial modes} \simeq \frac{\text{area } s_o/R^2}{\lambda^2/A} + 1 = \frac{(A)(\text{area } s_o)}{\lambda^2 R^2} + 1 \tag{47}$$

Note that the above interpretation effectively replaces the true source by a fictitious source whose spatial area corresponds to the area of the irradiance function $B(\underline{u})$.

B. Circular Detectors

Now consider the case where a circular detector of radius a is used, and assume a coherence-separable field as in part (A) having a coherence and radiance function possessing circular symmetry. That is, in (38)

$$\beta(\underline{r}_2-\underline{r}_1) = \beta(|\underline{r}_2-\underline{r}_1|) = \beta(r)$$
$$B(u) = B(|u|) \tag{48}$$

When the detector radius a is larger than the width of the coherence function, an approximate solution to (40) is given by

$$\hat{h}_{Km}(\underline{r}) = C_{Km}J_m(b_{Km}r)\cos m\theta_r \tag{49}$$

where C_{Km} are normalizing constants, $\{b_{Km}a\}$ are the zeros of the Bessel function $J_m(x)$, and θ is the angle of the point at distance r. Thus, (49) is the circular equivalent to (41). To determine eigenvalues we substitute back into the integral equation in (40). The right hand side becomes

$$\int_A \beta(\underline{r}-\underline{s})\ \hat{h}_{Km}(\underline{s})d\underline{s} = [A\beta(o)]^{-1} \int_A \beta(\underline{r}-\underline{s})\ C_{Km}J_m(b_{Km}s)\ \cos(m\theta_s)d\underline{s} \qquad (50)$$

where $[a\beta(o)]^{-1}$ is absorbed into C_{Km}. Using the transform identify in (38) in polar coordinates yields

$$\beta(\underline{r}-\underline{s}) = \int_0^\infty \rho B\left(\frac{\lambda R\rho}{2\pi}\right) J_o\left(\rho|\underline{r}-\underline{s}|\right)d\rho/2\pi$$

$$= \sum_{q=0}^{\infty} (2-\delta_{qo}) \int_0^\infty \rho B\left(\frac{\lambda R\rho}{2\pi}\right) J_q(\rho r)J_q(\rho s)\ \cos\ m(\theta_r-\theta_s)d\rho d\theta_s \qquad (51)$$

and integrating out θ_s, allows us to rewrite (50) as

$$C_{Km}\cos\ m\theta_r \int_0^\infty \rho B\left(\frac{\lambda R\rho}{2\pi}\right) J_m(\rho r)d\rho\ F_{Km}(\rho) \qquad (52)$$

with

$$F_{Km}(\rho) = \int_0^a sJ_m(\rho s)J_m(b_{Km}s)ds \qquad (53)$$

The term in (53) is approximately $\delta(\rho-b_{Km})/\rho$, so that (52) is

$$C_{Km}\cos(m\theta_r)J_m(b_{Km}r)\ B(\lambda Rb_{Km}/2\pi) \qquad (54)$$

This identifies the eigenvalue in (40) as

$$\gamma_{Km} = B(\lambda R b_{Km}/2\pi)$$

$$= B(\lambda R X_{Km}/2\pi a) \qquad (55)$$

where X_{Km} is the k^{th} zero of $J_m(X)$.

Thus, the eigenvalues are obtained by sampling the circular radiance function at distance $X_{Km}\lambda R/2\pi a$. The number of significant spatial modes could be identified similar to the previous example if the radiance area is taken as πa_o^2. Then the angle subtended by it at the source, when viewed from the detector is $\pi a_o^2/R^2$. Since each resolution area in (55) occupies a solid angle of $\lambda^2/\pi a^2$, the number of distinguishable modes is given by

$$\text{number modes} \cong \frac{\pi a_o^2/R^2}{\lambda^2/\pi a^2} + 1$$

$$= \frac{(\text{area radiance})(\text{area detector})}{\lambda^2 R^2} + 1 \qquad (56)$$

The result is the same as (47), with the rectangular areas replaced by the equivalent circular areas.

AN EXAMPLE

WHITE BANDLIMITED, RADIANCE LIMITED, GAUSSIAN FIELDS

The results of the previous section can now be applied to the most practical example - that in which the coherence-separable Gaussian field can be considered to have a flat power spectrum over a bandwidth B, and a flat radiance function over a spatial area a_o (projected normal to the detector line of sight). Let the detector itself have area A and observation time T, located R units from the source. For this case, the number of temporal modes is known to be $2BT + 1$, where all modes have identical eigenvalues γ_t. The number of spatial modes is $M + 1$, where $M = Aa_o/\lambda^2R^2$, and each has identical spatial eigenvalues γ_s (since samples of the radiance function are now all equal). This means the observed field, over a detector area and observation time T, has a total of $(2BT + 1)(M + 1)$ independent modes, each of eigenvalue $\gamma_s\gamma_t$. The resulting characteristic function in (27) is a product of a finite number of identical functions, and has the form

$$\psi_m(\omega) = \left[\frac{1}{1-\beta\gamma\omega}\right]^N \exp\left[\frac{N\beta E\omega}{1-\beta\gamma\omega}\right] \tag{57}$$

where

$$\gamma = \gamma_s\gamma_t$$

$$N = (2\beta T+1)(M+1)$$

Substitution into (26) can now be easily transformed yielding

$$p(k,V) = \frac{\gamma^k}{(1+\gamma)^{k+N}} \exp\left[-\frac{E}{1+\gamma}\right] L_k^N\left[-\frac{E}{\gamma(1+\gamma)}\right] \tag{58}$$

where E is the total signal energy over all N modes and $p(k,V)$ is the probability of k counts occurring over the volume $A \times T$. Note that the resulting density depends only upon the total signal energy E, the individual eigenvalues γ, and the total number of modes N. Thus, the primary result of extending the point detector to an extended spatial detector is to increase the effective number of modes over the given observation time T.

References

[1] S. Karp, E. O'Niell, and R. Gagliardi, "Communication Theory for the Free Space Optical Channel", Proc. IEEE Vol. 58, p. 1611, October 1970.

[2] S. Karp and J. Clark, "Photon Counting: A Problem in Noise Theory", Trans. on Information Theory Vol. IT-16, No. 6, p. 672, November 1970.

M-ary Poisson Detection and Optical Communications

ROBERT M. GAGLIARDI, MEMBER, IEEE, AND SHERMAN KARP, MEMBER, IEEE

REFERENCE: Gagliardi, R. M., and *Karp, S.: *M*-ARY POISSON DETECTION AND OPTICAL COMMUNICATIONS, University of Southern California, Los Angeles, Calif. 90007, and NASA Electronics Research Center, Cambridge, Mass. 02138. *Formerly with the University of Southern California. Rec'd 2/19/68; revised 9/3/68 and 12/20/68. Paper 69TP6-COM, approved by the IEEE Communication Theory Committee for publication without oral presentation. IEEE TRANS. ON COMMUNICATION TECHNOLOGY, 17-2, April 1969, pp. 208–216.

ABSTRACT: This paper presents an investigation of the problem of maximum likelihood detection of one of *M* Poisson processes in a background of additive Poisson noise. When the observables correspond to counts of emitted photoelectrons, the problem models a discrete version of a coherent *M*-ary optical communication system using photon counters in the presence of background radiation. Consideration is given to an average distance and a detection probability criterion. The advantages of an *M*-ary pulsed intensity set (Poisson intensities wholly concentrated in a single counting interval) are demonstrated. The performance of such intensity sets is exhibited in terms of error probabilities, pulse widths, signal-to-noise ratio, and channel capacity. Behavior as a function of number *M* of intensities is also discussed. By appropriate conversion these results may be used for determining power requirements in an optical pulse position modulation system.

I. Introduction

THE APPLICATION of detection theory to optical communications has been a subject of increasing interest. Since the output of a photodetecting surface is often modeled as sequences of electron "counts," and since optical photoelectrons have been generally accepted as obeying Poisson statistics, the analysis problem is basically one of signal detection involving Poisson processes. The problem was first formulated in this context by Reiffen and Sherman [1], and further contributions were made by Abend [2], Kailath [3], and Helstrom [4]. In this paper we investigate the general problem of *M*-ary detection based upon observations of events described by a time discrete Poisson process. Though the formulation of the problem is of a general nature, the principal application is to optical communications, and the practical limits of such a system will govern the mathematical assumptions imposed. Consideration is given to the divergence criterion for detection and to a criterion of maximization of probability of detection, both readily accepted as suitable design objectives. The

intensity set yielding optimal performance on the basis of special cases of these criteria is shown to be a special type of orthogonal intensity set, composed of M disjoint intensities wholly concentrated in a counting interval. Previously, the superiority of this type of signal set in binary detection had been shown by Abend [2] using a signal-to-noise ratio criterion, and by Kailath [3] using a distance criterion. This paper represents an extension of these results to M-ary Poisson detection.

The formulation of the problem follows that of Reiffen and Sherman [1]. The occurrence of events over an observed interval ΔT is said to obey a Poisson process if the probability of exactly k (an integer) events occurring is given by

$$P(k) = \frac{(n\Delta T)^k}{k!} e^{-n\Delta T}. \tag{1}$$

The parameter n is the average rate of occurrence and is called the intensity of the process. The average number of events occurring is then $n\Delta T$ and is often called the level of the process. If the events occur over a sequence of intervals ΔT in which the intensity may vary from one interval to the next, but is constant over each interval, we have a discrete time-varying Poisson process. In photodetection, each event corresponds to the emission of an electron, which occurs upon arrival of a photon, each photon having a fixed energy. The level is therefore proportional to the average energy received per interval, while the intensity n is proportional to the average power (see Section V). Thus, constraints upon level and intensity in Poisson processes are equivalent to energy and power constraints on the incident radiation.

In optical pulse-code modulation (PCM) communications information is transmitted, as shown in Fig. 1(a), by sending an optical signal intensity modulated with one of a set of possible intensities. The modulated signal is corrupted by background radiation of fixed intensity during reception, resulting in a process whose intensity is the sum of both intensities. The output of the photodetecting surface at the receiver is then a time-varying Poisson process of electron counts having the received intensity. In an M-ary system, the transmitter selects one of a set of M intensities for the optical process. The receiver, after photodetection, counts the number of electrons in each of M intervals ΔT seconds long and attempts to maximum likelihood detect which of M intensities is controlling the observed process. We shall assume ΔT is suitably shorter than the inverse bandwidth of the intensities so that the intensity remains approximately constant over ΔT. In addition, we assume that the counting interval is exactly known at the receiver by a perfect synchronization link. Thus, the above system can be modeled by the block diagram in Fig. 1(b). The input signal corresponds to a discrete Poisson process, while the interference appears as additive Poisson noise. (Recall that the sum of independent Poisson processes is itself a Poisson process having an intensity equal to the sum of the intensities.) The model

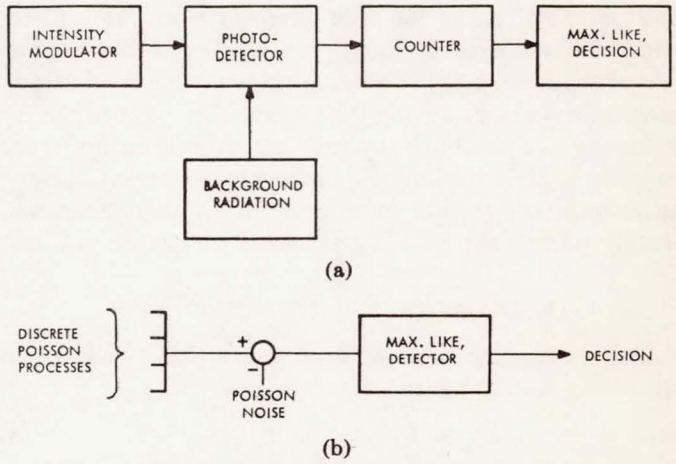


Fig. 1. A PCM optical communications receiver and its equivalent model

is idealized since other sources of interference, such as thermal noise and dark currents, are neglected. With this model the M-ary Poisson detection problem can be formulated as follows. Let a sequence of events obeying a discrete Poisson process occur over a sequence of M disjoint intervals ΔT, where $M\Delta T = T$, and the count over each interval is independent of all others. Let the observed process be controlled by one of M possible[1] intensity vectors $\boldsymbol{n}_q + \boldsymbol{n}_0$ for $q = 1,2,\cdots,M$, where

$$\begin{aligned} \boldsymbol{n}_q &= \{n_{q1}, n_{q2}, \cdots, n_{qM}\} \\ \boldsymbol{n}_0 &= \{n_0, n_0, n_0, \cdots, n_0\} \end{aligned} \tag{2}$$

for $n_{qi}, n_0 \geq 0$. The nonnegative n_{qi} is thus the intensity of $\boldsymbol{n}_q$ during the ith interval. Under a fixed energy constraint for each signal, we require

$$\sum_{i=1}^{M} n_{qi}\Delta T = N, \quad \text{for all } q. \tag{3}$$

The intensity vector $\boldsymbol{n}_0$ represents background noise of constant intensity superimposed upon the desired intensity. Let the corresponding number of events occurring in the ith interval be k_i. The problem then is to determine which of the possible intensity vectors $\boldsymbol{n}_q$ is controlling the received Poisson process by observing the sequence of independent counts $k = \{k_1, k_2, k_3, \cdots, k_M\}$. Under a maximum likelihood detection criterion and a priori equilikely intensities, it is well known [1] that the optimal test is to form the likelihood functions

$$\Lambda_q(k) = \sum_{i=1}^{M} \alpha_{qi} k_i \tag{4}$$

where

$$\alpha_{qi} = \log\left[\left(1 + \frac{n_{qi}}{n_0}\right)\right] \tag{5}$$

[1] In the statement of the problem, we assume M signals over M counting intervals. Subsequent discussion with the divergence criterion disproves the need for more than M intervals. The problem of designing M signals over fewer than M intervals is not considered here.

and select n_q+n_0 as the true intensity if no other likelihood function exceeds $\Lambda_q(k)$. If a likelihood draw occurs (more than one $\Lambda_q(k)$ is maximum), it is known that any randomized choice among the maxima can be used. In the following, we shall use a purely random selection in the case of likelihood draws. Equation (4) can be interpreted as a cross correlation of k with the α_{q_i}, an operation readily performed by a digital cross correlator [1].

II. Divergence of Detection Test

The divergence between two intensities n_j and n_q of the above test is defined as

$$D_{j_q} = E_{k/j}(\Lambda_{j_q}) - E_{k/q}(\Lambda_{j_q}) \tag{6}$$

where

$$\Lambda_{j_q} = \Lambda_j(k) - \Lambda_q(k)$$

and $E_{k/j}(\Lambda)$ is the conditional average of Λ with respect to k given the intensity n_j. Abend [2] has shown that for $M=2$ (binary detection) and the condition $n_2=0$, the divergence, normalized by the variance of Λ, is maximized by a "pulsed" type of intensity, where the level of the process is wholly concentrated in a single counting interval. Kailath [3] has extended this result by showing that, under a total energy constraint, other suitable forms of "distance" are maximized by similar pulsed intensities. We extend these notions here to the M-ary case and the equal energy constraint of (3).

The average divergence of an M-ary intensity set $\{n_q\}$ will be defined as

$$\overline{D} = \frac{1}{M^2} \sum_j \sum_q D_{j_q}. \tag{7}$$

Since $E_{k/j}(k_i) = (n_{j_i}+n_0)\Delta T$, the average divergence becomes

$$\begin{aligned}\overline{D} &= \frac{\Delta T}{M^2} \sum_j \sum_q \sum_i (n_{j_i} - n_{q_i}) \left\{ \log \left[1 + \frac{n_{j_i}}{n_0} \right] \right. \\ &\quad \left. - \log \left[1 + \frac{n_{q_i}}{n_0} \right] \right\} \\ &= \frac{2\Delta T}{M^2} \sum_i \left[M \sum_j n_{j_i} \log \left(1 + \frac{n_{j_i}\Delta T}{K} \right) \right. \\ &\quad \left. - \sum_{j \neq q} \sum n_{j_i} \log \left(1 + \frac{n_{q_i}\Delta T}{K} \right) \right] \end{aligned} \tag{8}$$

where $K = n_0\Delta T$. The nonnegativeness of the n_{j_i} and n_0 allows us to write

$$\begin{aligned}\overline{D} &\leq \frac{2\Delta T}{M} \sum_i \sum_j n_{j_i} \log \left(1 + \frac{n_{j_i}\Delta T}{K} \right) \\ &\leq \frac{2}{M} \left[\max_{i,j} \log \left(1 + \frac{n_{j_i}\Delta T}{K} \right) \right] \sum_j \sum_i n_{j_i}\Delta T \\ &\leq 2N \log \left(1 + \frac{N}{K} \right) \end{aligned} \tag{9}$$

as an upper bound under the constraint of (3). However, the first equality holds if the second term in (8) is zero, requiring n_{j_i}, for $j \neq q$, to be zero for all i at which n_{q_i} is nonzero. That is, the intensities must be mutually disjoint. The second and third equalities in (9) hold if $n_{j_i}=N$ for one i and $n_{j_i}=0$ for all other i. Thus, the upper bound for $\overline{D}$ occurs if the intensities of the set are disjoint and wholly concentrated in a single counting interval. This is satisfied with the set

$$n_q = \left\{ \frac{N}{\Delta T} \delta_{i_q} \right\}, \quad q = 1,2,\cdots,M \tag{10}$$

where δ_{i_q} is the Kronecker delta. The above represents an M-ary pulsed intensity set with each intensity occupying one of M intervals. It is significant that any disjoint intensity set, no matter how many intervals are used, yields the bound of the first inequality of (9), but only the pulsed intensity set of (10) yields the second bound. Thus, of all disjoint intensity sets, only the pulsed set maximizes $\overline{D}$, which immediately implies that only M intervals are required for maximization with M intensities. Last, it may be noted that with an average energy constraint over all intensities,

$$\frac{1}{M} \sum_j \sum_i n_{j_i}\Delta T = N \tag{11}$$

instead of (3), we have $n_{j_i} \leq MN/\Delta T$, and (9) becomes

$$\overline{D} \leq 2N \log \left(1 + \frac{MN}{K} \right) \tag{12}$$

which exceeds that previously derived. Furthermore, the upper bound in (12) occurs when $M-1$ intensities are zero everywhere, and one intensity is a pulsed intensity having value $MN/\Delta T$. In binary communications, for example, this means that an ON–OFF binary signal is superior to pulse position ($M=2$) signaling using the same average energy.

III. Detection Probability

The optimality of the M-ary pulsed intensity set has been shown, based on a divergence criterion. In this section we show that in certain cases this superiority also extends over a criterion based on maximization of the detection probability. We first require an expression for the detection probability for a general intensity set $\{n_q\}$. Usually this is obtained by first writing the conditional probability density of $\Lambda_q(k)$, then integrating over regions of correct decisioning. However, $\Lambda_q(k)$ in (4) is a weighted sum of independent Poisson variates which in general is not a Poisson variable. Rather, the true density involves an M-fold convolution of modified Poisson densities, yielding a result that is difficult to integrate. We shall instead use an alternative expression for the detection probability, derived in the Appendix, having the form

$$P_D = \frac{e^{-N}}{M} \sum_{R^M} \max_q \{\Psi(q,j)\} \tag{13}$$

where N is the intensity energy constraint of (3), R^M the space of all M-dimensional vectors j having nonnegative integer components, and

$$\Psi(q,j) = \prod_{i=1}^{M} \frac{[(n_{q_i} + n_0)\Delta T]^{j_i}}{j_i!} e^{-Mn_0\Delta T}. \tag{14}$$

The derivation of (13) follows an analogous procedure used in Gaussian channels (see [5]), but is somewhat complicated by the fact that likelihood draws occur with nonzero probability.

We would like to determine the intensity set $\{n_q\}$ for which P_D is maximum. This has been obtained for two particular cases of interest.

Case I: $M=2$ and Symmetric Intensity Sets

Let $M=2$ and consider the set of all possible symmetric intensity sets; i.e., if $n_1 = \{a,b\}$, then $n_2=\{b,a\}$. For this case it is easy to show that for any intensity set of this type, the vectors j for which $\Psi(1,j) \gtrless \Psi(2,j)$ when $a>b$, is simply the set $j=\{j_1,j_2\}$, such that $j_1 \gtrless j_2$. Using the constraint of (3) and letting $n_1=\{a,N-a\}$ and $n_2 = \{N-a,a\}$, for $N/2<a\leq N$, the detection probability is

$$P_D = \frac{e^{-(N+K)}}{2}\left\{\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{j_1-1} \frac{(a+K)^{j_1}}{j_1!}\cdot\frac{(N-a+K)^{j_2}}{j_2!} + \sum_{j_1=0}^{\infty}\sum_{j_2=j_1}^{\infty} \frac{(N-a+K)^{j_1}}{j_1!}\cdot\frac{(a+K)^{j_2}}{j_2!}\right\} \tag{15}$$

where again $K=n_0\Delta T$. Differentiating with respect to a yields

$$\frac{dP_D}{da} = \frac{e^{-(N+K)}}{2}\left\{\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{j_1-1} \frac{A^{j_1-1}}{(j_1-1)!}\cdot\frac{B^{j_2}}{j_2!} - \frac{A^{j_1}}{j_1!}\cdot\frac{B^{j_2-1}}{(j_2-1)!} + \sum_{j_1=0}^{\infty}\sum_{j_2=j_1}^{\infty} \frac{B^{j_1}}{j_1!}\cdot\frac{A^{j_2-1}}{(j_2-1)!} - \frac{A^{j_2}}{j_2!}\cdot\frac{B^{j_1-1}}{(j_1-1)!}\right\}$$

$$= \frac{e^{-(N+K)}}{2}\left\{1 + \sum_{j_1=1}^{\infty}\sum_{j_2=j_1-1}^{j_1} \frac{A^{j_1}B^{j_2} + B^{j_1}A^{j_2}}{j_1!j_2!}\right\}$$

where $A=(a+K)$ and $B=(N-a+K)$. Since A and B are positive, the above substantiates P_D as a monotone increasing function of a. Therefore P_D is maximized with a having its maximum value $a=N$ corresponding to the pulsed intensity set of (10) with $M=2$.

Case II: Any M, $N/K\to 0$

The limit above implies a high background noise level situation. We observe here that

$$P_D = \frac{e^{-N}}{M}\sum_{R^M}\max_q \Psi(q,j)$$

$$= \frac{e^{-N}}{M}\sum_{R^M} C \max_q\left\{\exp\left[\sum_{i=1}^{M} j_i \ln\left(1+\frac{n_{q_i}\Delta T}{K}\right)\right]\right\}$$

$$\xrightarrow[N/K\to 0]{} \frac{e^{-N}}{M}\sum_{R^M} C \max_q\left\{\exp\sum_{i=1}^{M}\frac{j_i n_{q_i}\Delta T}{K}\right\} \tag{16}$$

where

$$C = e^{-MK}\prod_i (1/j_i!)$$

and the limit follows since $n_{q_i}\Delta T/K \leq N/K \to 0$.

Now, with the constraint of (3),

$$\sum_i \frac{j_i n_{q_i}\Delta T}{K} \leq j_{\max}\frac{N}{K} \tag{17}$$

where $j_{\max} = \max_i\{j_i\}$. Thus

$$P_D \leq \frac{e^{-N}}{M}\sum_{R^M} C \exp\left\{j_{\max}\frac{N}{K}\right\}, \quad \frac{N}{K}\to 0.$$

The upper bound occurs when

$$\max_q\left\{\exp\sum_{i=1}^{M}\frac{j_i n_{q_i}\Delta T}{K}\right\} = \exp\left\{j_{\max}\frac{N}{K}\right\} \tag{18}$$

which clearly is true for the pulsed intensity set of (10), signifying asymptotic optimality for any M.

To determine optimal intensity sets (either global or local) in the general case, using (13), still remains a difficult task. It has been conjectured by many (e.g., see [1] and [3]) that the pulsed intensity set is in fact the optimal set, but to the authors' knowledge a rigorous proof has not been shown.

IV. Error Probabilities with Pulsed Intensity Sets

In this section we investigate the performance of the pulsed intensity set in M-ary detection by evaluating the error probability $P_E=1-P_D$. This can be obtained by using (13), but the computation can be more conveniently handled by noting that for the pulsed intensity set of (10) $\{\Lambda_q\}$ of (4) constitutes a set of independent Poisson random variables. The variables Λ_q have level $(N+K)$ if the qth intensity is sent, and have level K otherwise ($K=n_0\Delta T$). Recall that if the qth intensity is sent, a correct decision will be made with probability $1/(r+1)$ if Λ_q equals r other Λ's, and exceeds the remaining $M-1-r$. Therefore, upon considering all possibilities, the conditional detection probability is

$$P_{D/q} = \frac{e^{-(N+MK)}}{M} + \sum_{r=0}^{M-1}\sum_{x=1}^{\infty}\left[\frac{(N+K)^x}{x!}e^{-(N+K)}\right] \cdot\left[\sum_{t=0}^{x-1}\frac{K^t}{t!}e^{-K}\right]^{M-1-r}\left[\frac{K^x}{x!}e^{-K}\right]^r \cdot\left[\frac{(M-1)!}{r!(M-1-r)!(r+1)}\right]. \tag{19}$$

The right side is independent of q and thus represents the average detection probability. By applying the identity

$$\sum_{r=0}^{M-1}\frac{(M-1)!}{(r+1)!(M-1-r)!}A^{M-1-r}B^r = \frac{A^{M-1}}{M(B/A)}\left[\left(1+\frac{B}{A}\right)^M - 1\right]$$

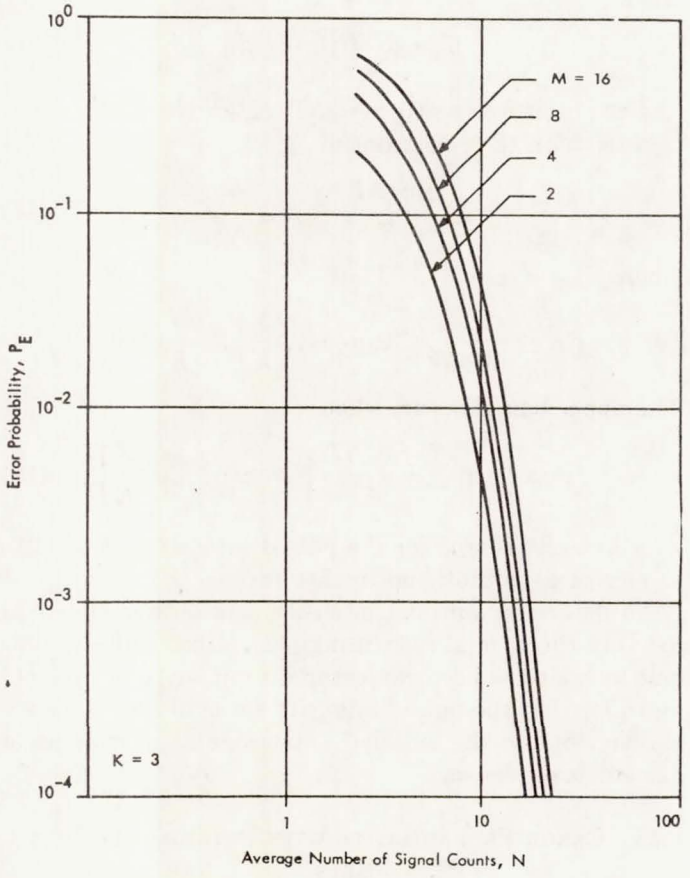


Fig. 2. Error probability versus normalized signal energy N for M-ary communications and $K=3$.

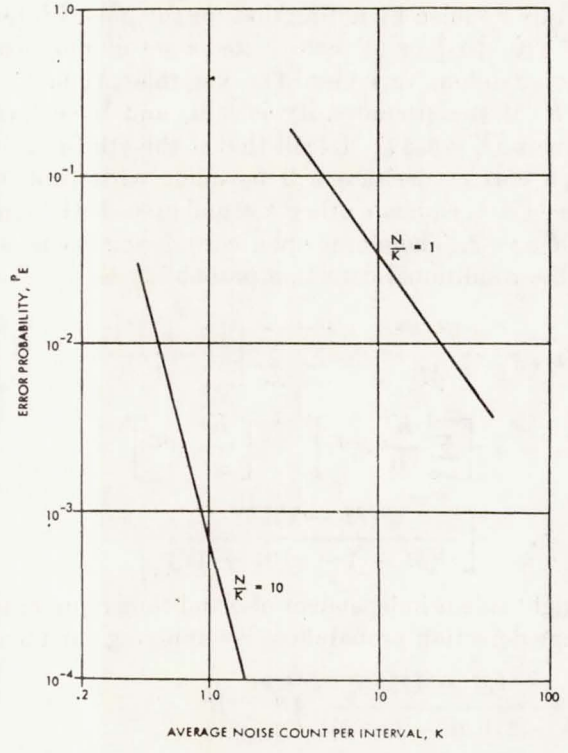


Fig. 3. Error probability versus normalized noise energy K. N = normalized signal energy.

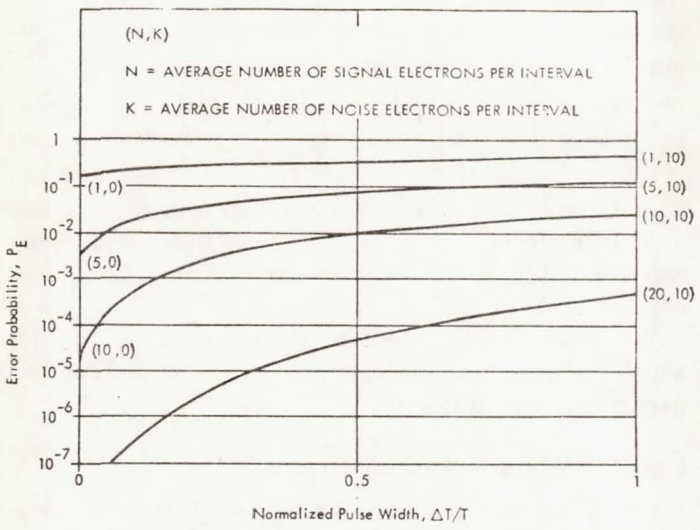


Fig. 4. Error probability versus normalized pulse width with $M=2$ and various operating conditions.

we can rewrite the error probability as

$$P_E(N,K,M) = 1 - P_D$$

$$= 1 - \frac{e^{-(N+MK)}}{M} - \sum_{x=1}^{\infty} \left[\frac{(N+K)^x e^{-(N+K)}}{x!}\right] \cdot \left[\sum_{t=0}^{x-1} \frac{K^t e^{-K}}{t!}\right]^{M-1} \frac{1}{Ma} [(1+a)^M - 1] \quad (20)$$

where

$$a = \frac{K^x}{x!\left[\sum_{t=0}^{x-1} \frac{K^t}{t!}\right]}.$$

The parameter $P_E(N,K,M)$ has been digitally computed for various values of N, K, and M. An exemplary plot is shown in Fig. 2 in which $P_E(N,3,M)$ has been plotted for various M as a function of N.

It is important to note that P_E depends on both the normalized signal energy N and the normalized noise energy K in the counting interval, and not simply on their ratio. This fact is emphasized in Fig. 3 in which $P_E(N,K,2)$ is plotted as a function of K for 2 fixed ratios N/K. This dependence on both signal and noise energies distinguishes the Poisson detection problem from the analogous coherent Gaussian channel problem. Note that the interfering noise energy K depends only upon the background energy in the interval ΔT, which is the width of the transmitted intensity pulse. The prime advantage of optical systems is precisely their ability to remove the effect of background noise by making ΔT small, and has been emphasized in previous reportings [6], [7]. This fact can be illustrated graphically, using (20), by considering a binary Poisson channel ($M=2$) sending information at

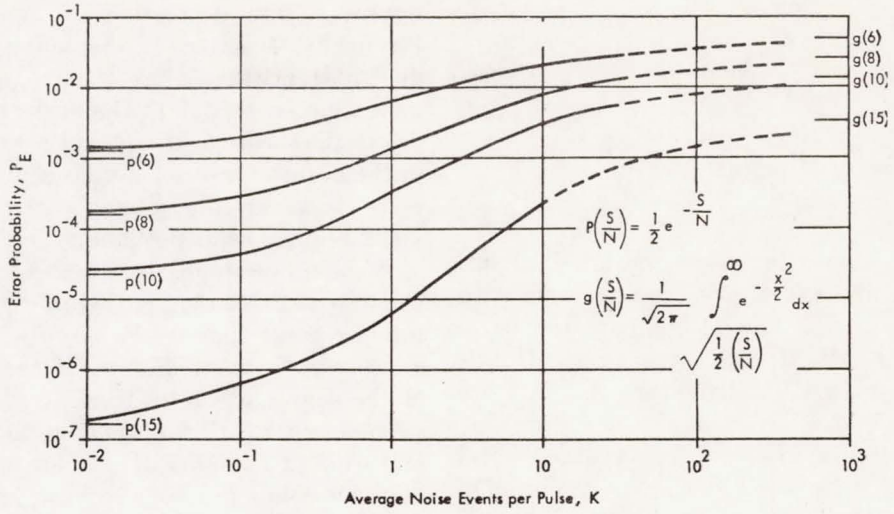


Fig. 5. Error probability versus normalized noise energy K for fixed values of $S/N = N^2/N + K$. $M = 2$.

a rate $1/T$ bit/s. The effect of the parameter ΔT is indicated by plotting $P_E(N, n_0 T \Delta T/T, 2)$ as a function of $\Delta T/T$, for fixed energy N and background noise energy per bit interval $n_0 T$. This is shown in Fig. 4. The results indicate the continuous improvement obtained by decreasing the "duty cycle" $\Delta T/T$, the ultimate limit corresponding to $\Delta T = 0$. The improvement, of course, is made at the expense of information bandwidth and peak power (both inversely proportional to ΔT). Surprisingly, the improvement is quite small at low values of N, and the increase in bandwidth may not be worth the decrease obtained in error probability.

A quantity of particular interest to communication engineers is the detected signal-to-noise ratio. This is often defined [8] as the ratio of the square of the average electron count with no noise to the variance of the count when noise is present. For Poisson counting statistics with pulsed intensities, this becomes $S/N = N^2/N + K$. The behavior of P_E of (20) as a function of K for fixed S/N is illustrated in Fig. 5 for a binary system. The results again indicate the ambiguity in using S/N as a design criterion. The asymptotes show the wide functional variation of P_E as K increases from zero.

As illustrated in Fig. 2, the error probabilities increase as M increases. However, the use of a single set of curves to compare various M-ary systems is misleading. An M-ary system with ΔT-second counting intervals transmits $\log_2 M$ bits of information in $M\Delta T$ seconds. It therefore communicates at a rate

$$R = \frac{\log_2 M}{M\Delta T} \quad \text{bit/s.} \tag{21}$$

If the transmission rate is normalized for each M, ΔT must be readjusted to maintain a fixed rate $R = R_0$. The effective noise level per counting interval is then

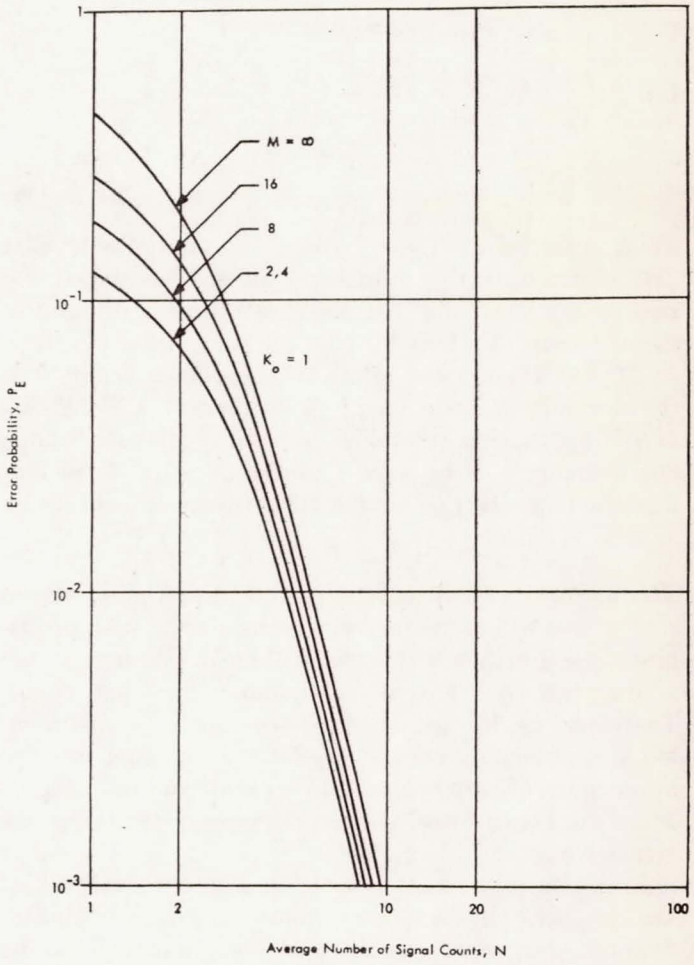


Fig. 6. Error probability versus normalized signal energy N, each M adjusted for fixed information rate R_0. K_0 = normalized noise energy per interval $1/2R_0$.

$$n_0\Delta T = n_0 \frac{\log_2 M}{MR_0} = \left(\frac{n_0}{2R_0}\right)\frac{2\log_2 M}{M} = 2K_0\frac{\log_2 M}{M} \tag{22}$$

where K_0 is the noise energy in an interval $1/2R_0$. Thus, for a comparison of different M-ary systems, each with fixed information rates, one should compare the parameter $P_E(N, 2K_0(\log M)/M, M)$ for each N. If this adjustment is made using (20), the curves of Fig. 6 are generated, with $K_0 = 1$.

The curve corresponding to $M = \infty$ is also shown, and is determined by taking the limit of $P_E(N, 2K_0(\log M)/M, M)$ as $M \to \infty$. This can be obtained by replacing K in (20) by $K' = 2K_0(\log M)/M$ and noting

$$\lim_{M\to\infty} e^{-(N+MK')} \to 0$$

$$\lim_{M\to\infty}\left[\sum_{t=0}^{x-1}\frac{(K')^t}{t!}e^{-K'}\right]^{M-1}\left[\frac{(1+a)^M-1}{Ma}\right] \to \begin{cases} 0, & x = 1 \\ 1, & x > 1. \end{cases}$$

Using the above we then have

$$\lim_{M\to\infty} P_E\left(N, 2K_0\frac{\log M}{M}, M\right) \to 1 - \sum_{x=2}^{\infty}\frac{(N)^x e^{-N}}{x!} = 1 - (1 - Ne^{-N} - e^{-N}) = (1+N)e^{-N} \tag{23}$$

which is plotted as $M = \infty$ in Fig. 6. It is noteworthy that (23) is precisely the probability of an event count of zero or one occurring in a noiseless counting interval of signal energy N. This has the following interpretation. As $M \to \infty$, the number of intervals becomes infinite, but the normalized noise energy per interval $K' = 2K_0(\log M)/M$ approaches zero. The probability that more than one event will occur in any one of $M-1$ independent nonsignaling intervals having noise energy K' is given by

$$1 - [(1 + K')e^{-K'}]^{M-1}$$

This approaches zero as $M \to \infty$, indicating that counts of zero or one will occur in every such interval with probability one. Furthermore, there will be an infinite number of intervals with a zero count and with a one count. Therefore, as $M \to \infty$, an error will occur (with probability approaching one) whenever the signaling interval has a count of zero or one, and an error will never occur when the latter interval has a count greater than one. Hence we have (23).

It is also interesting to note in Fig. 6 that the best system operation, in terms of minimal error probability, does not always correspond to $M \to \infty$. In fact, it can be shown that best M operation depends strongly on the amount of background noise K_0. For example, if $K_0 \approx 0$, it is easy to show, using (20), that for M finite, $P_E(N, K', M) \approx (M-1)e^{-N}/M$, which is monotonically decreasing with M and always less than the $M = \infty$ value of (23). Thus, with negligible background noise, system operation improves with decreasing M, and is best for $M = 2$. Physically, this means the noise reduction advantage due to decreasing ΔT as M increases does not offset the increasing errors due to the larger numbers of likelihood draws that will occur. (Recall a random choice is made in the event of draws.) For large amounts of background noise, however, the converse is true, and $M = \infty$ does yield minimal error probability.

It should be emphasized that a fixed energy constraint was imposed on the signal intensity, and therefore the time average power $P_0 = N/T = NR/\log M$ actually approaches zero as $M \to \infty$. If the average power level of the source has been fixed at some level P_0, N must be replaced by $P_0 \log M/R$ in the previous equations, and we find $P_E \to 0$ is $M \to \infty$ for any $P_0 > 0$. This result may be compared to a similar result for an additive Gaussian channel [9] in which zero error probability occurred only if P_0 satisfied a condition dependent on the rate R.

The P_E results above are useful for determining the channel capacity (maximum information rate) of an M-ary pulsed intensity set. Assume a transmitter sends one of an M pulsed intensities every T seconds, with each pulse having width $\Delta T = T/M$. If the transmitter operates at a fixed rate R_0, then again $T = (\log M)/R_0$ as given by (21). The channel can now be represented as a symmetric channel in which each of the M equally likely intensities is converted to itself with probability $1 - P_E$, and is converted to any of the other intensities with equal probability $P_E/(M-1)$. The channel capacity for this type of system is known to be

$$C = \frac{\log M + P_E \log (P_E/(M-1)) + (1-P_E)\log(1-P_E)}{\log M/R_0} \tag{24}$$

where $P_E = P_E(N, n_0 \log M/MR_0, M)$. We shall again consider the signal intensity energy N, the background noise power n_0, and the rate R_0 to be held fixed. Then, as $M \to \infty$, P_E approaches the limit in (23), while the channel capacity has the limit

$$C \to [1 - (1+N)e^{-N}]R_0 \tag{25}$$

for N finite. The above indicates that information transfer can be forced to approach any desired rate with a finite signal energy by using an increasingly larger number of intensities and adjusting R_0 at the transmitter. However, each level is transmitted with a nonzero error probability, and the information bandwidth and peak power become infinite. Again introduction of a transmitter power constraint, instead of an energy constraint, will yield operation at a capacity R_0 with a zero error probability as $M \to \infty$.

V. Summary and Application of Results

In this paper we have investigated M-ary Poisson detection, defined as the maximum likelihood detection of one of a set of M discrete Poisson processes in the presence of an additive discrete Poisson noise process.

The model represents a discrete version of an optical communication system in which the observables are counts of photoelectrons, the signals are intensity modulated continuous-wave optical sources, and the noise is background radiation received within the optical bandwidth. The photoelectron count can then be modeled as a time varying Poisson process whose average rate is proportional to the sum of the intensities of the modulated source and the background radiation. In practical operation, the intensity of the optical source is a continuous process, but the analysis can be put on a discrete basis by partitioning the signaling intervals into subintervals over which the intensity is taken to be constant. The above Poisson model is examined, and the advantages of a pulsed type of intensity set are demonstrated. The latter corresponds to an optical system using pulse position modulation in which information is transmitted by a burst or pulse of optical energy located in one of a set of pulse positions. The performance of such a system, in terms of pulse width and numbers of pulse positions, is presented here. The results of this paper basically represent theoretical limits which an optical link can approach, since the deleterious effects of receiver (thermal) noise have been neglected. This latter assumption becomes valid, for example, when photomultipliers are used in detection, and the background radiation collected at the receiver is the predominant source of noise.

The analyses and performance results are in terms of N and K, the average electron counts due to signal and noise, respectively. However, these results can be easily converted to average power requirements by using the relations

$$N = \eta P_s M/hfB$$

$$K = \eta P_n/hfB$$

where P_s and P_n are the average signal and background noise power, $h = 6.62 \times 10^{-34}$ J·s, η is the photodetector efficiency (including photomultiplication), f is the optical frequency of the continuous-wave source, and $B = 1/\Delta T$. The average power P_s and P_n can be further converted to transmitted power by introducing space losses and receiver optics (e.g., see [10, chs. 1 and 2]). Exact synchronization has been assumed here between transmitter and receiver at all times. Besides receiver thermal noise, the analysis has excluded the effects of photomultiplier statistics, saturation, and dark currents. We also have assumed constant intensity background radiation which implies a wide-band optical filter. This assumption restricts the minimum value of ΔT to approximately 10^{-10}–10^{-12} second.

Appendix

In this Appendix we derive (13) of the report. The average probability of correctly determining the true intensity in M-ary transmission is

$$P_D = \frac{1}{M} \sum_{q=1}^{M} P(D|q) \tag{26}$$

where $P(D|q)$ is the probability of correct detection, given that $n_q + n_0$ is the true intensity. Now the conditional probability of the occurrence of an observed vector $k = j = \{j_1, j_2, j_3, \cdots, j_M\}$, given $n_q + n_0$, is

$$P(k = j|q) = \prod_{i=1}^{M} \frac{[(n_{q_i} + n_0)\Delta T]^{j_i}}{j_i!} e^{-(Mn_0)\Delta T} e^{-N}$$

$$\triangleq \Psi(q,j)\, e^{-N} \tag{27}$$

where N is the energy constraint given in (3). The conditional detection probability $P(D|q)$ is then obtained by summing over the set of all j, such that a correct decision is made. A correct decision will occur when the qth intensity is used, if Λ_q is selected as being the largest. If no other Λ_t exceeds Λ_q, but r of the Λ_t equal Λ_q, the receiver will be correct with a probability of $1/(r+1)$, assuming a purely random selection is made when likelihood equalities occur. Now j is an M dimensional vector with nonnegative integer components, and we shall denote the space of all such vectors as R^M. The conditional detection probability $P(D|q)$ can therefore be written by summing over all $j \in R^M$ leading to a correct decision. Thus,

$$P(D|q) = \sum_{r=0}^{M-1} \frac{1}{r+1} \sum_{J_{qr}} \Psi(q,j)\, e^{-N} \tag{28}$$

where J_{q_r} is the set of $j \in R^M$ such that no other Λ_t exceeds Λ_q, and r other Λ_t equal Λ_q. If we let I_q denote the r dimensional index set corresponding to these r Λ_t, we can for simplicity denote J_{q_r} symbolically as

$$J_{q_r} = \{j \subset R^M: \ \Lambda_q = \max_k \Lambda_k = \Lambda_t, t \subset I_q\}. \tag{29}$$

Substituting (28) into (26) yields a general expression for the detection probability:

$$P_D = \frac{e^{-N}}{M} \sum_{q=1}^{M} \sum_{r=0}^{M-1} \sum_{J_{qr}} \frac{1}{r+1} \Psi(q,j). \tag{30}$$

Now by examining carefully the set J_{q_r} we can simplify the above. Making use of the monotonicity of the exponential function, we can write:

$$\begin{aligned} J_{q_r} &= \{j \subset R^M: \ \exp(\Lambda_q) = \exp(\max_k \Lambda_k) \\ &= \exp \Lambda_t, t \subset I_q\} \\ &= \Big\{j \subset R^M: \ \prod_{i=1}^{M} [(n_{q_i} + n_0)\Delta T]^{j_i} \\ &= \max_k \prod_{i=1}^{M} [(n_{k_i} + n_0)\Delta T]^{j_i} \\ &= \prod_{i=1}^{M} [(n_{t_i} + n_0)\Delta T]^{j_i}, \quad t \subset I_q\Big\} \\ &= \{j \subset R^M: \ \Psi(q,j) = \max_k \Psi(k,j) \\ &= \Psi(t,j), t \subset I_q\}. \end{aligned} \tag{31}$$

Thus J_{q_r} can be alternatively defined as the set of j for which $(\Psi q,j)$ is one of $r+1$ maximum $\Psi(k,j)$ functions. This means every j in J_{q_r} also belongs to r other

sets J_{tr}, $t \subset I_q$, or correspondingly, a point j in J_{tr}, $t \subset I_q$, exists such that $\Psi(t,j) = \Psi(q,j)$. Note that the set of subspaces $\{J_{qr}\}$ are disjoint for different r, but not for different q. With these facts consider the summation

$$\sum_{q=1}^{M} \sum_{J_{qr}} \frac{\Psi(q,j)}{r+1} \tag{32}$$

for fixed r. For any term of the sum, say $\Psi(q_0,j_0)/(r+1)$, there exist r other terms having the same value, one for each point j_0 of J_{tr}, $t \subset I_{q0}$. The total contribution to the sum above from this set of $r+1$ terms is then

$$(r+1)\left[\frac{\Psi(q_0,j_0)}{r+1}\right] = \Psi(q_0,j_0) = \max_q \Psi(q,j_0) \tag{33}$$

the last equation following since $j_0 \in J_{q0r}$. Thus, overlapping points in the summation of (32) contribute a total amount given by (33). It therefore follows that

$$\sum_{q=1}^{M} \sum_{J_{qr}} \frac{\Psi(q,j)}{r+1} = \sum_{\bigcup_q J_{qr}} \max_q \Psi(q,j) \tag{34}$$

where $\bigcup_q J_{qr}$ is the union over q of the subsets $\{J_{qr}\}$. Inverting the order of summation in (30) and using (34) allows us to rewrite

$$P_D = \frac{e^{-N}}{M} \sum_{r=0}^{M-1} \sum_{\bigcup_q J_{qr}} \max_q \Psi(q,j) = \frac{e^{-N}}{M} \sum_{RM} \max_q \Psi(q,j) \tag{35}$$

where we have employed the fact that the $\bigcup_q J_{qr}$ are disjoint subspaces, and the sum over all r spans the whole space R^M. Equation (35) is the same as (13).

References

[1] B. Reiffen and H. Sherman, "On optimum demodulator for Poisson processes: photon source detectors," *Proc. IEEE*, vol. 51, pp. 1316–1320, October 1963.

[2] K. Abend, "Optimum photon detection," *IEEE Trans. Information Theory (Correspondence)*, vol. IT-12, pp. 64–65, January 1966.

[3] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Communication Technology*, vol. COM-15, pp. 52–60, February 1967.

[4] C. W. Helstrom, "The detection and resolution of optical signals," *IEEE Trans. Information Theory*, vol. IT-10, pp. 275–287, October 1964.

[5] A. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966, p. 234.

[6] M. Ross, "Pulse interval modulation laser communications," presented at the Eastcon Convention, Washington, D. C., October 1967.

[7] S. Karp and R. Gagliardi, "A low duty cycle optical communication system," presented at the Eastcon Convention, Washington, D. C., October 1967.

[8] W. Pratt, "Binary detection in an optical polarization modulation communication channel," *IEEE Trans. Communication Technology (Concise Papers)*, vol. COM-14, pp. 664–665, October 1966.

[9] A. Viterbi [5], p. 226.

[10] M. Ross, *Laser Receivers*. New York: Wiley, 1966.

Robert M. Gagliardi (S'57–M'61) was born in [redacted], on [redacted]. He received the B.S. degree in electrical engineering from the University of Connecticut, Storrs, in 1956, and the M.S. and Ph.D. degrees in engineering from Yale University, New Haven, in 1957 and 1960, respectively.

From 1958 to 1960 he was an Instructor at the New Haven Engineering College. In 1960 he joined the Information Study Section, Space System Division, Hughes Aircraft Company, Culver City, Calif., where he was involved in problems in telemetry and communication systems. He is presently an Associate Professor in the Department of Electrical Engineering, University of Southern California, Los Angeles, and a Consultant to Hughes Aircraft Company.

Dr. Gagliardi is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

Sherman Karp (M'62) was born in [redacted] on [redacted]. He received the B.S.E.E. and M.S.E.E. degrees from the Massachusetts Institute of Technology, Cambridge, in 1960 and 1962, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1967.

After several years of industrial experience, he joined NASA Electronics Research Center, Cambridge, Mass. He is presently Section Chief in Optical Communications in the Optics Laboratory. His main interest is in the general area of reliable communication systems with interest in modulation and coding techniques.

Dr. Karp is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

NASA supported under Grant NGL 05-018-104

Reprinted by permission from
IEEE TRANSACTIONS ON COMMUNICATIONS
Vol. COM-22, No. 5, May 1974

Estimation of Delay of M PPM Signals in Laguerre Communications

N. C. MOHANTY, MEMBER, IEEE

***Abstract*—In this note the maximum likelihood detector for M pulse-position-modulated (PPM) signals in Laguerre communications is derived. A decision-directed maximum likelihood estimator for the delay of M PPM signals is discussed.**

INTRODUCTION

The output of an idealized quantum photo detector is a doubly stochastic Poisson process with the intensity of the process

$$\lambda(t) = \beta \mid S(t) + n(t) \mid^2$$

where $n(t), t \epsilon (0,T)$, is zero-mean white Gaussian band limited to $\pm B$ Hz and $s(t)$ is a deterministic signal occupying the same band. Karp and Clark [1] have shown that the probability of the number of counts N_T in $[0,T]$ is K and is given by Laguerre distribution, i.e.,

$$P[N_T = K] = \frac{(\beta N_0)^K}{(1+\beta N_0)^{1+K+\frac{1}{2}BT}} \cdot \exp\left[-\beta \int_0^T \mid S(t) \mid^2 dt \Big/ (1+\beta N_0)\right] \cdot L_K^{\frac{1}{2}BT}\left[-\int_0^T \mid S(t) \mid^2 dt \Big/ N_0(1+\beta N_0)\right] \quad (1)$$

where β is a constant directly proportional to the detector quantum efficiency and surface area and inversely proportional to photon energy. For a wide range of signal intensities $\mid S(t) \mid^2$, this can be approximated by a Poisson distribution with intensity

$$\beta \mid S(t) \mid^2 + \beta \bar{P}$$

where $\beta N_0 \ll 1$ and $2BT \gg 1$. $\bar{P}$ is the total average noise power $2BN_0$. Estimation of the delay of one signal in optical communication is discussed by Karp and Clark [1] and Bar-David [2]. In this note we want to estimate the delay of M pulse-position-modulated (PPM) signals in Laguerre communications when the receiver does not know which signal is present. In this M-ary PPM system, the transmitter selects one of the set of M intensities for the optical process. The receiver, after photo detection, counts the K_1, the number of electrons in each $(M+1)$ intervals, ΔT seconds long, and attempts to maximum likelihood detect which of M intensities is controlling the observed process. On the basis of this decision, the maximum likelihood estimate of the delay θ, $0 < \theta < \Delta T$, is derived when the signal-to-noise ratio (SNR) is very high and it is compared with distribution of Poisson statistics.

MAXIMUM LIKELIHOOD DETECTOR

Assuming the cost matrix $\{c_{ij}\}$ is such that $c_{ii} = 0$, $c_{ij} = 1$, $i \neq j$, and the signals are equilikely, we decide S_i if joint distribution $(K_1, \cdots, K_{N+1})$ satisfies

$$P(K_1, K_2, \cdots, K_{M+1}/S_i) = \max_{1 \leq j \leq M} P(K_1, K_2, \cdots, K_{M+1}/S_j). \quad (2)$$

Since our assumption is white band-limited noise, counts in each interval, i.e., K_i counts in the interval $((i-1)\Delta T, i\Delta T)$, are independent

$$P(K_1, K_2, \cdots, K_{M+1}/S_j) = \prod_{i=1}^{M+1} P(K_i/S_j).$$

Signals are given by

$$S_i(t) = \begin{cases} \sqrt{E}, & \text{if } \Delta T(i-1) < t < i\Delta T \\ 0, & \text{elsewhere.} \end{cases}$$

Since (1) is generated by convolution of M-fold, the joint distribution of $K_1, K_2, \cdots, K_{M+1}$ is given by

$$P(K_1, K_2, \cdots, K_{M+1}/S_j) = a^K b L_{K_i}((\Delta T - \theta)c) L_{K_{i+1}}(-\theta c)$$

where

$$a = \frac{\beta N_0}{1+\beta N_0}, \qquad K_t = K_1 + K_2 + \cdots + K_{M+1}$$

$$b = \frac{1}{1+\beta N_0} \exp\left[-\frac{\Delta T E}{1+\beta N_0}\right]$$

$$c = \frac{E}{N_0(1+\beta N_0)}, \qquad \theta = \text{delay parameter.}$$

Therefore, (2) becomes

$$L_{K_i}((\Delta T - \theta)c) L_{K_{i+1}}(-\theta c) = \max_{1 \leq j \leq M} L_{K_j}(-(\Delta T - \theta)c) L_{K_{j+1}}(-\theta c)$$

then s_i is sent. Since this maximization is true for all θ, and $L_K(-x)$ is a monotonically increasing function in $x \geq 0$, the rule is to decide s_i if

$$\int L_{K_i}(-(\Delta T - \theta)c L_{K_{i+1}}(-\theta c)\, d\theta = \max_{1 \leq j \leq M} \int L_{K_j}(-(\Delta T - \theta)c) L_{K_{j+1}}(-\theta c)\, d\theta.$$

Using identities [3], and after some algebra, decide s_i if

$$L_{K_i + K_{i+1}}{}^{+1}(-\Delta TE) = \max_{1 \leq j \leq M} L_{K_j + K_{j+1}}{}^{+1}(-\Delta TE).$$

Since Laguerre polynomials are increasing, the decision criteria becomes decide s_i if

$$K_i + K_{i+1} = \max_{1 \leq j \leq M} K_j + K_{j+1}.$$

Paper approved by the Associate Editor for Space Communications of the IEEE Communications Society for publication without oral presentation. Manuscript received March 2, 1973; revised November 23, 1973.

The author is with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, N. Y.

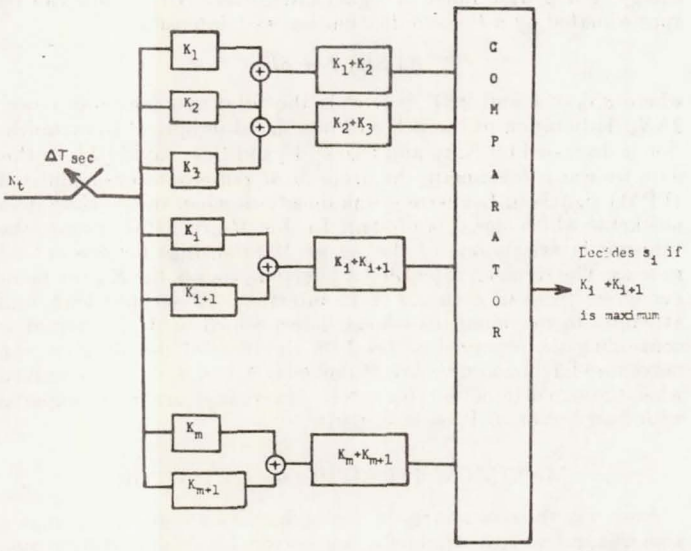


Fig. 1. Maximum likelihood detector of M PPM signals.

The receiver counts the number of photons in each of two adjacent counters and compares and decides s_i if the count is maximum in the ith counter of $2\Delta T$. See Fig. 1. When the counting statistics are Poisson (i.e., when the SNR is very high), it can be easily verified that the decision criteria is the same as the preceding statements.

MAXIMUM LIKELIHOOD ESTIMATOR

Based on the decision that s_i is present, we will find that the maximum likelihood estimator $\hat{\theta}_L$·$\hat{\theta}_L$ must satisfy

$$\frac{d}{d\theta}[L_{K_i}[-(\Delta T-\theta)c]L_{K_{i+1}}(-\theta c)]=0. \tag{3}$$

Again using Laguerre identities [3], (3) becomes

$$L_{K_{i-1}}{}'(-\Delta T-\hat{\theta}_L)c)L_{K_{i+1}}(-\hat{\theta}_L c) = L_{K_i}(-(\Delta T-\hat{\theta}_L)c)L_{K_{i+1-1}}{}'(-\hat{\theta}_L c).$$

Assuming very high SNR,

$$L^a(x)\sim\frac{(-x)^K}{K!}, \qquad \text{when } x\gg 1. \tag{4}$$

Using (3) and (4), $\hat{\theta}$ must satisfy

$$+\frac{(\Delta T-\theta_L)c}{K_j}=+\frac{\hat{\theta}_L c}{K_{j+1}}. \tag{5}$$

From (5), we get

$$\hat{\theta}_L=\frac{\Delta T K_{j+1}}{K_j+K_{j+1}}.$$

When the counting statistics are Poisson, the maximum likelihood estimator for $\hat{\theta}$ must satisfy

$$\frac{d}{d\theta}[[(\Delta T-\hat{\theta}_P)E+\bar{P}]^{K_i}[\hat{\theta}_P E+\bar{P}]^{K_{i+1}}]=0.$$

This, on simplification, gives

$$\frac{(\Delta T-\hat{\theta}_P)E+\bar{P}}{K_i}=\frac{\hat{\theta}_P E+\bar{P}}{K_{i+1}}. \tag{6}$$

From (6), we get

$$\hat{\theta}_P=\frac{\Delta T E K_{i+1}+\bar{P}(K_{i+1}-K_i)}{E(K_i+K_{i+1})}.$$

When $E/P\gg 1$, $\hat{\theta}_P$ and $\hat{\theta}_L$ are almost equal.

CONCLUSION

An adaptive estimator based on decision criteria is derived on the assumption of very high SNR. A minimum mean-square estimator can be derived if we assume some statistics on θ. Related matters are discussed in [4]–[6].

ACKNOWLEDGMENT

The author wishes to thank Prof. R. M. Gagliardi for helpful discussions.

REFERENCES

[1] S. Karp and J. R. Clark, "Photon counting: A problem in classical noise theory," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 672–680, Nov. 1970.
[2] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 31–37, Jan. 1969.
[3] G. Sansone, *Orthogonal Functions*. New York: Interscience, 1959, ch. 4.
[4] R. Gagliardi and N. Mohanty, "Estimation theory and optical communication," Electron. Sci. Lab., Univ. Southern California, Los Angeles, USCEE Rep. 446, Apr. 1973.
[5] ——, "MAP synchronization in optical communication systems," Electron. Sci. Lab., Univ. Southern California, Los Angeles, USCEE Rep. 448, Apr. 1973.
[6] N. C. Mohanty, "M-ary Laguerre detection," *IEEE Trans. Aerosp. Electron. Syst.* (Corresp.), vol. AES-9, pp. 464–467, May 1973.

Reprinted by permission from IEEE TRANSACTIONS ON INFORMATION THEORY
Vol. IT-18, No. 4, July 1972, pp. 514-515

On the Identifiability of Finite Mixtures of Laguerre Distributions

N. C. MOHANTY

***Abstract*—Finite mixtures of Laguerre distribution are identifiable.**

Introduction

Identifiability is a necessary criterion for estimating parameters in a mixture distribution. It is known [1], [2] that the classes of normal, exponential, Cauchy, and negative binomial distributions are identifiable. It was established by Feller [3] that mixtures of Poisson distributions are identifiable. In this correspondence we show that mixtures of Laguerre distributions are identifiable using a theorem of Teicher [4]. It is worthwhile to mention that the Poisson distribution is a limiting case of the Laguerre distribution, both of which are encountered in optical communication theory [5].

Laguerre and Poisson Mixtures

We shall use the following theorem.

Theorem [4]: Let $\mathcal{F} = \{F\}$ be a family of cumulative distributions with transforms $\Phi(t)$ defined for $t \in S_\Phi$ (the domain of definition of Φ) such that the mapping M: $F \to \Phi$ is linear and one to one. Suppose that there exists a total ordering ($\leq$) of $\mathcal{F}$ such that $F_1 < F_2$ implies i) $S_{\Phi_1} \leq S_{\Phi_2}$, ii) the existence of some $t_1 \in \bar{S}_{\Phi_1}$ (t_1 being independent of Φ_2) such that $\lim_{t \to t_1} \Phi_2(t)/\Phi_1(t) = 0$. Then the class K' of all finite mixtures of $\mathcal{F}$ is identifiable.

Proposition: The class of all finite mixtures of Laguerre distributions is identifiable.

Manuscript received December 10, 1971; revised February 15, 1972. This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-014.

The author is with the Department of Electrical Engineering, University of Southern California, Los Angeles, Calif. 90007.

Proof:

$$P(N = K) = \exp\left(\frac{xZ}{1 - Z}\right) Z^K (1 - Z)^{\alpha+1} L_K^\alpha(x),$$

where $L_K^\alpha(\ )$ is the associated Laguerre polynomial of order α, x is the parameter, and $|Z| < 1$.

The Laplace transform of $P(N = K)$ [6] is denoted by

$$\Phi(t) = \frac{(1 - Z)^{\alpha+1}}{(1 - eZ^{-t})^{\alpha+1}} \exp\left[x\left(\lambda - \frac{Ze^{-t}}{1 - Z^{-t}}\right)\right],$$

where $\lambda = Z/(1 - Z)$. If $x_2 > x_1$, then $P(x_1,K) < P(x_2,K)$

$$\frac{\Phi_2(t)}{\Phi_1(t)} = \exp\left[(x_2 - x_1)\left(\lambda - \frac{Ze^{-t}}{1 - Ze^{-t}}\right)\right].$$

Now with $S_\Phi = (-\infty, +\infty)$ and $t_1 = \log Z$

$$\lim_{t \to t_1} \frac{\Phi_2(t)}{\Phi_1(t)} = 0.$$

Conclusion

When the counting statistics are Poisson or Laguerre, finite mixtures of the counting statistics are identifiable.

References

[1] S. J. Yakowitz, "Unsupervised learning and identification of finite mixtures," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 330–338, May 1970.
[2] S. J. Yakowitz and J. Spragins, "On the identifiability of finite mixtures," *Ann. Math. Statist.*, vol. 39, pp. 209–214, Feb. 1968.
[3] W. Feller, "On a general class of contagious distributions," *Ann. Math. Statist.*, vol. 14, pp. 389–433, 1943.
[4] H. Teicher, "Identifiability of finite mixtures," *Ann. Math. Statist.*, vol. 34, pp. 1265–1269, 1963.
[5] S. Karp and J. R. Clark, "Photon counting: A problem in classical noise theory," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 672–680, Nov. 1970.
[6] N. C. Mohanty, "Synchronization for adaptive optical communications," Ph.D. dissertation, Univ. Southern California, Los Angeles, 1972.

M-ary Laguerre Detection

Abstract

In this correspondence, maximum-likelihood *M*-ary detection theory is applied to an incoherent optical system model employing photo detectors governed by Laguerre counting statistics. It is shown that a maximum-likelihood Laguerre detector corresponds to a count comparison over each signaling interval. Error probability is derived.

I. Introduction

In a communication system using optical devices, the receiver is modeled as a counter of electrons which are emitted from a photodetector when a modulated optical wave in the form of photons is incident on it. The synthesis of the optimal receiver processing and its resulting performance depend upon the statistics associated with this counting. The probability of obtaining K photoelectrons in the finite time interval $(0, t)$ due to a general radiant source is Poisson distributed according to the function

Manuscript received June 5, 1972.
This work was sponsored under Contract NGR-05-018-104.

$$P[U_{R,t} = K|I(t)] = \frac{1}{K!}\left[\int_0^t o' I(t')\,dt\right]^K \cdot \exp\left[-\alpha'\int_0^t I(t')\,dt'\right]$$

where α' is constant associated with the photodetector, $U_{R,t}$ is the random variable that represents the number of photoelectron counts, and $I_{(t)}$ is the intensity of the light. When $I_{(t)}$ is a random function of time, the sum of a deterministic signal and a random noise, the probability distribution is a random function of time. The required stationary counting distribution $P(U_{R,t} = K)$, which characterizes the detection process, is found by the statistical average of $P[U_{R,t}|I(t)]$:

$$P[U_{R,t} = K] = \frac{1}{K!}\int_0^\infty \left[\alpha'\int_0^t I(t')\,dt\right]^K \cdot \exp\left[-\alpha'\int_0^t I(t')\,dt'\right] P(I)\,dt \quad (1)$$

where $P(I)$ is the probability distribution of the intensity $I(t)$ of the optical wave incident upon the detector [7, pp. 11-12]. In general, $P[U_{R,t} = K] \triangleq P(K)$ is a double stochastic Poisson process [8]. We will consider the case in which intensity is the complex envelope of the optical field; i.e.,

$$I(t) = |S(t) + n(t)|^2$$

where $S(t)$ is a deterministic signal and $n(t)$ is zero-mean Gaussian noise band-limited to $\pm B$ Hz. With $\alpha' = 1$ and signal bandwidth less than noise bandwidth, taking first, $(2Bt + 1)$ eigenvalues of the covariance of $n(t)$, it can be shown [1] that the probability distribution of K electrons is given by

$$P(K) = \frac{N^K}{(1+N)^{\alpha+K+1}} \exp\left[\frac{-E}{1+N}\right] L_K^\alpha\left[\frac{-E}{N(1+N)}\right] \quad (2)$$

where N is the average number of noise photon counts per time space mode, given by Planck's formula $N = [\exp(hf/e\mu) - 1]^{-1}$, where h is Planck's constant, e is Boltzmann's constant, and f is the optical carrier frequency, E is the energy of the signal,

$$L_K^\alpha(x) = \sum_{m=0}^{K}\binom{K+\alpha}{K-m}\frac{(-x)^m}{m!},$$

and $\alpha = 2Bt$ is the time-bandwidth product.

When the counting statistics are Poisson, the detection and estimation problems are considered by Reiffen and Sherman [2], Bar-David [3], and Gagliardi and Karp [4]. In this correspondence, we will consider detections of M signals when counting statistics are governed by Laguerre distribution. Detection schemes for two signals in this area are considered by Helstrom [5] and Gagliardi [6].

II. M-ary Laguerre Detection

In an M-ary system, the transmitter selects one of a set of M intensities for the optical process. The receiver, after photodetection, counts the number of electrons in each of M intervals $\Delta T (M\Delta T = t$ seconds long, and attempts maximum-likelihood detection of which of M intensities is controlling the observed process. We shall assume that ΔT is suitably shorter than the inverse bandwidth of the intensities, so that the intensity remains approximately constant over ΔT. In addition, we assume that the counting interval is exactly known at the receiver by a perfect synchronization link, signals are equilikely, and the cost matrix is given by C_{ij}, $c_{ii} = 0$, $c_{ij} = 1$, $i \neq j$. After observing $K = (K_1, K_2, K_3, \ldots, K_M)$, where K_i is the count in the interval $[(i-1)\Delta T, i\Delta T]$ and $M\Delta T = t$, we decide S_i if

$$P(K|S_i) = \max_{1\leq j\leq M} P(K|S_j). \quad (3)$$

Since our assumption is white noise, the K_i are independent,

$$P(K|S_j) = \prod_{i=1}^{M} P(K_i|S_j). \quad (4)$$

Signals are given by

$$S_i(t) = \begin{cases} E, & \text{if } \Delta T(i-1) \leq t \leq i\Delta T \\ 0, & \text{elsewhere.} \end{cases} \quad (5)$$

Since (2) is generated by convolution of M-fold

$$P(K_i|S) = \frac{N^{K_i}}{(1+N)^{K_i+1}} \exp\left[\frac{-E^i}{1+N}\right] L_{K_i}\left[\frac{E^i}{N(1+N)}\right]$$

where E^i is the signal energy in the ith counting interval, and from (5), we get

$$P(K_i|S_i) = \frac{N^{K_i}}{(1+N)^{K_i+1}} \exp\left[\frac{-E}{1+N}\right] L_{K_i}\left[\frac{-E}{N(1+N)}\right]$$

$$P(K_j|S_i) = \frac{N^{K_i}}{(1+N)^{1+K_i}}, \qquad i \neq j. \quad (6)$$

Using (4), (5), and (6),

$$P(K|S_j) = \left(\prod_{i=1}^{M} \frac{N^{K_i}}{(1+N)^{1+K_i}}\right) \exp\left[\frac{-E}{1+N}\right] \cdot L_{K_j}\left[\frac{-E}{N(1+N)}\right]. \quad (7)$$

Substituting (7) into (3) and cancelling common terms, we get

$$L_{K_i}\left[\frac{-E}{N(1+N)}\right] = \max_{1\leqslant j\leqslant M} L_{K_j}\left[\frac{-E}{N(1+N)}\right] \quad (8)$$

if S_i is decided. $L_m[-x]$ is a monotonic increasing function if $x \geqslant 0$. Therefore, $K_i = \max_{1\leqslant j\leqslant M} K_j$ if S_i is decided. The scheme is to make a comparison of counts, and then to pick the maximum one. When $M = 2$, our result agrees with Gagliardi [6].

III. Error Probability

Our decision scheme suggests that, after counting electrons in M counters, we select S_i if the count in the ith count is maximum. Following the procedure of Gagliardi and Karp [4], probability of error PE, when signal S_i is decided, is given by

$$\begin{aligned}
\mathrm{PE} = {} & \sum_{K_i} \sum_{K_{i-1}=K_i} \cdots \sum_{K_M=K_i} P(K_1|S_i)P(K_2|S_i)P(K_M|S_i) \\
& - \left\{\binom{M}{2} \sum_{K_i=0}^{\infty} \sum_{K_{i-1}=0}^{\infty} \sum_{K_{i-2}=K_i}^{\infty} \cdots \sum_{K_M=K_i}^{\infty} P(K_1|S_i) \right. \\
& \times P(K_2|S_i)P(K_M|S) \\
& + \binom{M}{3} \sum_{K_i=0}^{\infty} \sum_{K_{i-1}}^{\infty} \sum_{K_{i-2}}^{\infty} \sum_{K_{i-3}=K_i}^{\infty} \cdots \sum_{K_M=K_i}^{\infty} P(K_1|S_i) \\
& \cdot P(K_2|S_i)P(K_M|S) \\
& + \cdots \\
& + \cdots \\
& \left. + \binom{M}{M-1} \sum_{K_i=0}^{\infty} P(K_i|S_i)P(K_i|S_i)\cdots P(K_i|S_i) \right\}. \quad (9)
\end{aligned}$$

Using [7], we get, from (8),

$$\begin{aligned}
\mathrm{PE} = {} & \left[\frac{1}{1+N}\right]^M \left[\sum_{K_i} a^{K_i} b L_{K_i}[C] \frac{a^{(M-1)K_i}}{(1-a)^{M-1}} \right. \\
& - \left\{\frac{1}{2}\binom{M}{2}\left(\sum a^{K_i} b L_{K_i}[C]\right)\right. \\
& \cdot \left(\sum a^{K_i-1} b L_{K_i-1}[C] \frac{a^{(M-2)K_i-1}}{(1-a)^{M-2}}\right) \\
& + \frac{1}{3}\binom{M}{3}\left(\sum a^{K_i} b L_{K_i}[C]\right)\left(\sum a^{K_i-1} b L_{K_i-1}[C]\right) \\
& \cdot \left(\sum a^{K_i-2} b L_{K_i-2}[C] \frac{a^{(M-3)K_i}}{(1-a)^{M-1}}\right) \\
& + \cdots \\
& + \cdots \\
& \left. \left. + \binom{M}{M-1} \sum a^{K_i} b L_{K_i}[C]^M \right\}\right] \quad (10)
\end{aligned}$$

where

$$a = \frac{N}{1+N}, \quad b = \exp\left[-\frac{E}{1+N}\right]$$

$$C = \frac{-E}{N(1+N)}.$$

By using the identity

$$\sum_{K=0}^{\infty} Z^K L_K(x) = (1-z)^{-1} \exp\left[\frac{xz}{z-1}\right],$$

we get, from the (10),

$$\begin{aligned}
\mathrm{PE} = {} & \left(\frac{1}{1+N}\right)^M \left\{\frac{b}{(1-a)^{M-1}} (1-a^M)^{-1} \left[\exp \frac{-Ca^M}{a^M-1}\right] \right. \\
& - \left[\frac{1}{2} - \binom{M}{2} \frac{b^2}{(1-a)^{M-1}} (1-a^{M-1})^{-1} \exp\left[\frac{Ca^{M-2}}{a^{M-1}-1}\right] \right. \\
& \cdot (1-a)^{-1} \exp\left[\frac{-Ca}{a-1}\right] + \frac{1}{3}\binom{M}{3} \frac{b^3}{(1-a)^{M-3}} \\
& \cdot (1-a^{M-2}) \exp\left[\frac{Ca^{M-2}}{a^{M-2}-1}\right] (1-a)^{-2} \exp\left[\frac{-2Ca}{a-1}\right] \\
& \left. \left. + \frac{1}{M}\binom{M}{M+1} b^M \exp\left[\frac{-MCa}{a-1}\right]\right]\right\}.
\end{aligned}$$

On further simplification, we get

$$PE = \frac{1}{1+N}\left[\frac{b \exp\left[\dfrac{ca^M}{1-a^M}\right]}{1-a^M}\right] - \sum_{r=2}^{M} \frac{1}{r}\binom{M}{r-1}\frac{b^r}{(1-a^{M-r+1})} \cdot \left\{\exp\left[\frac{-rca}{a-1}\right] + \frac{ca^{M-r+1}}{a^{M-r+1}-1}\right\}.$$

This result can be computed very easily for various values of M, E, and N

Acknowledgment

The author is indebeted to Prof. R. M. Gagliardi for valuable discussions.

N. C. MOHANTY
University of Southern California
Los Angeles, Calif. 90007

References

[1] S. Karp and J.R. Clark, "Photon counting: A problem in classical noise theory," *IEEE Trans. Information Theory*, vol. IT-16, November 1970.

[2] B. Reiffen and H. Sherman, "On optimum demodulator for Poisson process: Photon source detector," *Proc. IEEE*, vol. 51, pp. 1316-1320, October 1963.

[3] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Information Theory*, vol. IT-15, January 1969.

[4] R.M. Gagliardi and D. Karp, "*M*-ary Poisson detection and optical communication," *IEEE Trans. Communication Technology*, vol. COM-17, April 1969.

[5] C.W. Helstrom, "Performance of an ideal quantum receiver of a coherent signal of random noise," *IEEE Trans. Aerospace and Electronic Systems*, pp. 562-566, May 1969.

[6] R.M. Gagliardi, "Photon counting and Laguerre detection," *IEEE Trans. Information Theory*, vol. IT-18, January 1972.

[7] W.K. Pratt, *Laser Communication Systems*. New York: Wiley, 1969.

[8] D.L. Snyder, "Filtering and detection for doubly stochastic Poisson process," *IEEE Trans. Information Theory*, vol. IT-18, pp. 91-101, January 1972.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 58, NO. 10, OCTOBER, 1970
pp. 1611–1626

Communication Theory for the Free-Space Optical Channel

S. KARP, MEMBER, IEEE, E. L. O'NEILL, AND R. M. GAGLIARDI, MEMBER, IEEE

Abstract—The current understanding of quantum detectors, the noise mechanisms which limit (are basic to) their operation, and their application to optical communications (theory) is summarized. In this context, we are considering channels in which the electromagnetic field is not subjected to any propagation effects other than a geometric loss. (Such a channel would exist between satellites.) Consequently, we will concentrate on optimum time processing using the tools of statistical communication theory.

Fundamental to the study of a detection process is the need to develop a good mathematical model to describe it [1]–[6]. Therefore, approximately one-fifth of the paper is devoted to establishing, in a semi-classical analysis, the quantum detector output electron number as a conditional Poisson process with the conditioning variable being the modulus of the electromagnetic field. Once this has been established, these results are used to derive various limiting probability densities related to actual practice. Although the mathematical details are omitted, these results will be presented from the viewpoint of orthogonal function expansions and interpreted in terms of an eigenspace.

The resulting current flow is analyzed next as a shot noise process, and the power density spectrum is calculated. Attention is focused on isolating the signal components from the noise in terms of both the current probability density and the power density spectrum. Examples are given where appropriate. At this point, an understanding of the underlying noise processes will have been presented and attention will shift to analog and digital communications.

The analog communication will be presented primarily in terms of the signal-to-noise ratio. The *S/N* ratio in direct detection will be presented both as a ratio of the integrals of two separate portions of the spectrum and as a ratio of two moments of the probability density describing the current. These calculations will be extended to include heterodyne detection.

Digital communications will be discussed in the context of detection theory. It will be shown that the likelihood ratio is often a monotonic function of the random variable representing the number of electrons flowing. Hence optimum processing will consist of a weighted count of electrons from various counting modes. Digital design will be presented in terms of *M*-ary signaling, error probabilities, and information rates.

Manuscript received May 15, 1970.

S. Karp was with the NASA Electronics Research Center, Cambridge, Mass. He is now with DOT Transportation Systems Center, Cambridge, Mass.

E. L. O'Neill is with the Worcester Polytechnic Institute, Worcester, Mass.

R. M. Gagliardi is with the University of Southern California, Los Angeles, Calif.

I. INTRODUCTION

WE BEGIN with a classical description for the energy and momentum densities of a radiation field for both the single- and multimode cases. Confining our treatment to the semi-classical theory, we sketch briefly the argument that the probability of ejecting an electron from a photo-cathode surface in a short time interval is proportional to the light intensity. From this point of view, we deduce an expression for the probability of releasing n photoelectrons in a time T in terms of a weighted Poisson distribution. The weight factor is the probability distribution for the accumulated energy received on the photodetective surface in equal times.

A. Semi-Classical Theory

1) Normal Mode Decomposition of the Field: We begin our description of the semi-classical theory of radiation and matter by writing down the free space wave equation for the vector potential $\vec{A}(\vec{r}, t)$

$$\nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = 0. \tag{1}$$

Electing to work in the Coulomb gauge, div $\vec{A}=0$, the electric and magnetic field vectors are now given by

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t}$$

$$\vec{B} = \text{curl } \vec{A}. \tag{2}$$

Concentrating first on a single mode of the radiation field, a plane wave is characterized by the components of the wave vector $\vec{k}=(k_x, k_y, k_z)$ where $\omega=|k|c$. However, even after specifying the direction and frequency of a plane electromagnetic wave, there still exists the possibility of two independent, orthogonal polarization directions, $\hat{\sigma}_1$ and $\hat{\sigma}_2$.

A plane wave, then, at frequency ω propagating in the direction $\hat{k}$ can be written as:

$$\vec{A}(\vec{r}, t)=\vec{a}(t)\exp(i\vec{k}\cdot\vec{r})+\vec{a}^*(t)\exp(-i\vec{k}\cdot\vec{r}) \tag{3a}$$

$$\vec{E} = i\omega(\vec{a}\exp(i\vec{k}\cdot\vec{r})-\vec{a}^*\exp(-i\vec{k}\cdot\vec{r})) \tag{3b}$$

$$\vec{B}=i[(\vec{k}\times\vec{a})\exp(i\vec{k}\cdot\vec{r})-(\vec{k}\times\vec{a}^*)\exp(-i\vec{k}\cdot\vec{r})] \tag{3c}$$

where

$$\vec{a} = (a_1\hat{\sigma}_1 + a_2\hat{\sigma}_2)\exp(-i\omega t).$$

It will also turn out to be useful to list the energy density μ plus the linear and angular momentum densities $\vec{g}$ and $\vec{m}$ associated with this wave.

$$\mu = \frac{\vec{E}\cdot\vec{D}+\vec{B}\cdot\vec{H}}{2} = 2\omega^2\varepsilon_0|a|^2$$

$$\vec{g} = \frac{\vec{E}\times\vec{H}}{c^2} = -\frac{1}{\mu_0 c^2}(\dot{\vec{A}}\times\text{curl }\vec{A}) = \frac{2\omega^2\varepsilon_0}{c}|a|^2\hat{k}$$

$$\vec{m} = \frac{1}{\mu_0 c^2}(\vec{A}\times\dot{\vec{A}}) = 2\omega\varepsilon_0(|b_+|^2 - |b_-|^2)\hat{k} \tag{4}$$

where

$$b_\pm = \frac{\sqrt{2}}{2}(a_1 \pm ia_2).$$

We are following here the notation of Louisell [7]. The ambiguity in sign in the last expression is removed when we choose either right- or left-handed circularly polarized light. Of course, for linearly polarized light, a_1 and a_2 are in phase so that with $|b_+|^2=|b_-|^2$ no net angular momentum is propagated. We also add in passing that the second term in (3a) is added to ensure the reality of $\vec{A}=\vec{A}^*$. A plane wave traveling in the opposite direction $(-\hat{k})$ is obtained by changing the sign of $\vec{k}$. Finally, a standing wave is described by taking a linear combination of the expression with $+\vec{k}$ and $-\vec{k}$. Before moving on to the multimode description of the radiation field, we will now select a single polarization component of the field and decompose this complex quantity in the form:

$$a_j = \frac{1}{\sqrt{4\varepsilon_0\omega^2}}(\omega q_j + ip_j)$$

$$a_j^* = \frac{1}{\sqrt{4\varepsilon_0\omega^2}}(\omega q_j - ip_j). \tag{5}$$

Under this transformation of variables, the energy and momentum densities become

$$\mu_j = \frac{p_j^2 + \omega^2 q_j^2}{2} = H_j$$

$$g_j = \frac{H_j}{c}\hat{k} \tag{6}$$

so that as far as energy and momentum considerations are concerned, the radiation field can be treated as a simple harmonic oscillator obeying Hamilton's canonical equations of motion

$$\dot{q}_j = \frac{\partial H_j}{\partial p_j}$$

$$\dot{p}_j = -\frac{\partial H_j}{\partial q_j}. \tag{7}$$

Turning now to the multimode description of the field, we impose periodic boundary conditions by introducing the triad of integers l_1, l_2, and l_3 into the relation

$$\vec{k} = (k_x, k_y, k_z) = \frac{2\pi}{L}(l_1, l_2, l_3). \tag{8}$$

For economy in notation, we will henceforth use the symbol l to imply this triad, and all Fourier sums will be treated as:

$$\sum_{l_1}\sum_{l_2}\sum_{l_3} \Rightarrow \sum_{l}. \tag{9}$$

Moreover, the orthogonality relation

$$\int_0^L\int_0^L\int_0^L \exp[i(\vec{k}_l - \vec{k}_{l'})\cdot\vec{r}]\,dx\,dy\,dx = V\delta_{ll'} \tag{10}$$

taken over a cube of volume $V=L^3$ will guarantee that each mode will contribute independently[1] to the total energy and momentum of the field.

We are now in position to put all these pieces together. Starting with the multimode description of the vector potential

$$\vec{A}(\vec{r}, t) = \sum_l\sum_\sigma a_{l\sigma}\exp(i\vec{k}_l\cdot\vec{r}) + \text{complex conjugate} \tag{11}$$

and introducing the canonical variables $q_{l\sigma}$ and $p_{l\sigma}$ through the relation

$$a_{l\sigma} = \frac{1}{\sqrt{4\varepsilon_0 V\omega_l^2}}(\omega_l q_{l\sigma} + ip_{l\sigma}) \tag{12}$$

we may now list the expressions for the total energy and momentum of the field in the form

$$U = \sum_l\sum_\sigma 2\varepsilon_0 V\omega_l^2 a_{l\sigma}^* a_{l\sigma} = \sum_l\sum_\sigma \frac{p_{lr}^2 + \omega_l^2 q_{lr}^2}{2}$$

$$= \sum_l\sum_\sigma H_{l\sigma} = H$$

[1] Of course, this lack of cross terms in adding up the total energy of a system is the whole idea behind normal mode decomposition. Also, choosing plane wave eigenfunctions, orthogonal over cubic geometry, is merely the simplest way to proceed. Ultimately, we will work with the mode density, in which case the size and shape of the cavity will not appear.

$$\vec{G} = \sum_{l,\sigma}\sum \frac{2\varepsilon_0 V\omega_l^2}{c} a_{l\sigma}^* a_{l\sigma}\hat{k}_l = \sum_{l,\sigma}\sum \frac{H_{l\sigma}}{c}\hat{k}_l. \tag{13}$$

These equations indicate that so far as energy and momentum are concerned, the radiation field may be considered as a collection of oscillators, each contributing (per mode) to the total energy and momentum. We point out here that a quantum oscillator's level of excitation is given by $H_{l\sigma} = n_{l\sigma}\hbar_{\omega l}$ and when this condition is inserted into (13) there results the conclusion that a radiation field may be treated as a superposition of discrete photons, $n_{l\sigma}$ per mode, each possessing energy $\hbar\omega_l$ momentum $\hbar\omega_l/c$ and angular momentum $\pm\hbar$.

2) Interaction Between an Atom and a Radiation Field: A complete description of the emission and absorption of light by an atom influenced by a radiation field is well beyond the scope of this paper. The reader, interested in the details of the process, is urged to consult [7]–[10]. We present here only a bare outline of the approach insofar as it related to the photon counting distribution.

Starting with the complete Hamiltonian for a charged particle in an electromagnetic field

$$H = \frac{(\vec{p} - e\vec{A})^2}{2m} + H_R + eV \tag{14}$$

we neglect the term in e^2 and use the gauge condition div $A = 0$ to reduce this to

$$H = H_A + H_R + H_I \tag{15}$$

where $H_A = (p^2/2m) + eV$ is the Hamiltonian of the atom,

$$H_R = \sum_l \sum_\sigma H_{l\sigma}$$

is the Hamiltonian of the radiation field, and $H_I = -(e/m)\vec{A}\cdot\vec{p}$ is the interaction Hamiltonian. Combining the first two terms into the unperturbed Hamiltonian $H_0 = H_A + H_R$, we next treat H_I as a perturbation and attempt to solve the Schrödinger equation

$$(H_0 + H_I)|\psi\rangle = i\hbar\frac{\partial|\psi\rangle}{\partial t}. \tag{16}$$

Using the method of first order perturbation theory, we attempt an expansion of $|\psi\rangle$ into a linear combination (with time varying coefficients) of the eigenstate $|\psi_n^0\rangle$ of the unperturbed Hamiltonian, known to satisfy the equation

$$H_0|\psi_n^0\rangle = i\hbar\frac{\partial|\psi_n^0\rangle}{\partial t}. \tag{17}$$

In this expansion we have

$$|\psi\rangle = \sum_n C_n(t)\exp\left(-\frac{i}{\hbar}E_n t\right)|\psi_n^0\rangle \tag{18}$$

and the probability of finding the system in a state $|\psi_n^0\rangle$ is

$$|C_n(t)|^2 = |\langle\psi_n^0|\psi\rangle|^2. \tag{19}$$

Assuming then that the combined system, atom plus radiation field, begins in some initial state, $|i\rangle$, (18) implies a set of coupled equations for the probability amplitude ($C_n(t)$) from which one can determine $|C_f(t)|^2$, the probability of finding the combined system in the final state $|f\rangle$. Summing over all final states, and making a number of simplifying assumptions [8]–[10], one ends up with Fermi's "golden rule" for the probability per second for a transition in the form

$$\frac{dP}{dt} \cong \frac{2\pi}{\hbar}|\langle f|H_I|i\rangle|^2\rho(E_f). \tag{20}$$

Here $\rho(E_f)$ is the density-in-energy of the final states, and $\langle f|H_I|i\rangle$ represents the matrix element of the perturbation Hamiltonian between the initial and final states. When applied to the problem of an atom in a radiation field, one must distinguish between the cases when only the atom and both the atom and the radiation field are treated as quantized systems. In the former, the semi-classical treatment, one can correctly deduce Einstein's B coefficient for stimulated emission and absorption in terms of the electric dipole moment taken between the initial and final wave functions of the atom. On the other hand, when one also quantizes the field including the zero point fluctuation, then (20) also predicts the existence of Einstein's A coefficient for spontaneous emission.

3) Photon Counting Statistics: The consequence of (20) that is of importance to us is that it leads [8] to the result that in a short time Δt the probability of ejecting an electron from an atom on the surface of a photocathode is proportional to the incident intensity of the light $I(t)$. That is

$$P_1(t, t + \Delta t) = \alpha I(t)\,\Delta t. \tag{21}$$

For sufficiently short times $P_0(t, t+\Delta t) \cong 1 - \alpha I(t)\,\Delta t$ so that in an interval $(0, t + \Delta t)$ there are but two ways of releasing n photo-electrons, given by

$$P_n(0, t + \Delta t) = P_{n-1}(0, t)\alpha I(t)\Delta t + P_n(0, t)(1 - \alpha I(t)\Delta t). \tag{22}$$

Subtracting $P_n(0, t)$ from both sides and dividing by Δt before passing to the limit, we can write

$$\frac{dP_n}{dt} = \alpha I(t)P_{n-1}(t) - \alpha I(t)P_n(t). \tag{23}$$

The solution to this differential-difference equation is

$$P_n(t) = \frac{\left[\alpha\int_0^t I(t')\,dt'\right]^n \exp\left[-\alpha\int_0^t I(t')\,dt'\right]}{n!}. \tag{24}$$

Now if this process were carried out a number of times over similarly prepared realizations of the field, the average over this ensemble would lead to

$$P_n(t, T) = \frac{\int_0^\infty (\alpha w)^n \exp(-\alpha w)}{n!} P(w)\,dw \tag{25}$$

where

$$w = \int_t^{t+T} I(t')\,dt'$$

and $P(w)dw$ is the probability for w to lie in the range $(w, w + dw)$.

4) Mode Density: So far as the question of density of radiation modes is concerned, we can start from one of several points of view. From the viewpoint of wave optics, light of wavelength λ emerging from a slit of width Δx can be expected to produce interference and diffraction effects over an angle $\Delta\alpha$ such that $\Delta x\Delta\alpha \sim \lambda$. Extending this notion to the elemental area $\Delta s = \Delta x\Delta y$ we see that

$$\Delta\alpha\Delta\beta \sim \frac{\lambda^2}{\Delta s} = \frac{\Delta A}{R^2}. \tag{26}$$

In terms of the "coherence area," this can be written as

$$\Delta A \sim \frac{\lambda^2}{\Delta s/R^2} = \frac{\lambda^2}{\Delta\Omega}. \tag{27}$$

Further, if light proceeding from Δs has a bandwidth $\Delta\nu$ then there exists a "coherence time" $\Delta t \sim 1/\Delta\nu$ corresponding to a "coherence length" $\Delta l = c\Delta t \sim c/\Delta\nu$. Dividing by two to take into account the two independent polarization states, we now write for the "coherence volume"

$$\Delta V = \frac{\Delta A\Delta l}{2} = \frac{(c\Delta t)\lambda^2}{2\Delta\Omega} = \frac{1}{2\Delta\Omega(\nu^2/c^3)\Delta\nu}. \tag{28}$$

In a volume V, we expect to find $\Delta N = V/\Delta V$ modes, or in terms of mode density

$$N_\nu = \frac{\Delta N}{V\Delta\nu} = (2)(\Delta\Omega)\frac{\nu^2}{c^3}. \tag{29}$$

For isotropic radiation, this reduces to the familiar expression

$$N_\nu = \frac{8\pi\nu^2}{c^3}. \tag{30}$$

From a purely quantum statistical point of view, the elementary cell size in phase space is given by

$$\Delta x\Delta p_x\Delta y\Delta p_y\Delta z\Delta p_z \sim h^3 \tag{31}$$

so that for a beam of photons of momentum $p = h\nu/c$ in a solid angle $\Delta\Omega$ about p we have:

$$\Delta x\Delta y\Delta z \sim \frac{h^3}{p^2\Delta p\Delta\Omega} = \frac{h^3}{(h^3\nu^2/c^3)\Delta\nu\Delta\Omega}. \tag{32}$$

Dividing by two to account for the two orthogonal polarization states, we end up with, again

$$\Delta V = \frac{1}{2\Delta\Omega(\nu^2/c^3)\Delta\nu}. \tag{33}$$

It is important therefore to know how many spatial and temporal modes of the radiation field interact with the photo-detector. We shall see that a single mode of chaotic thermal radiation, and stabilized laser radiation lead, respectively, to Bose-Einstein and Poisson photocount distributions. For the case of several radiation modes, one needs to calculate the probability distribution for the sum of several random variables leading to multiple convolutions.

II. Compound Photocurrent Distributions

It is clear in view of the preceding discussion that when using a quantum detector, one always has a Poisson process governing the current flow. That is, the number N_t of electrons flowing in any interval $(0, t)$ is a random variable. If the time-space envelope of the projected electromagnetic (EM) field $|V(\tau, \mathbf{r})|$, $(0, t)$ is given, then the probability density for $N_t = k$ electrons to flow in this interval is

$$P_{N_t}(k) = \frac{\left[\int_0^t \int_A \beta|V(\tau, \mathbf{r})|^2\, d\tau\, d\mathbf{r}\right]^k}{k!} \cdot \exp\left[-\int_0^t \int_A \beta|V(\tau, \mathbf{r})|^2\, d\tau\, d\mathbf{r}\right]. \tag{34}$$

If on the other hand the quantity $|V(\tau, \mathbf{r})|$ is random or has a random component, then (1) is a conditional density and must be written as $P_{N_t}(k/|V(\tau, \mathbf{r})|)$. To find $P_{N_t}(k)$ requires the additional averaging

$$P_{N_t}(k) = E_{|V(t,\mathbf{r})|}\left[P_{N_t}(k/|V(t, \mathbf{r})|\right]. \tag{35}$$

For the purpose of this discussion we will assume that the integration over the detector surface merely yields a constant (e.g., a point detector) and that we can write

$$\int_A \int_0^t \beta|V(\tau, \mathbf{r})|^2\, d\tau\, d\mathbf{r} = \alpha \int_0^t |a(\tau)|^2\, d\tau$$

with $\alpha = \eta/h\nu$, η the quantum efficiency, and $|a(\tau)|^2$ the instantaneous power in the received process. Notice that $|a(\tau)|$ is the envelope of the received process and that (35) really amounts to performing the final average over the statistics of the envelope.

In most communication problems (and the ones which we will consider), the function $a(\tau)$ can be expressed as the linear sum of a known signal $s(\tau)$ and a noise process $n(\tau)$. The signal may also contain a stochastic parameter σ to represent a channel disturbance such as fading. As is common at lower frequencies, the component $n(\tau)$ can be accurately modelled as a Gaussian noise process. Hence we will assume that $a(\tau)$ can be written as

$$a(\tau, \sigma) = s(\tau, \sigma) + n(\tau)$$

which is the complex envelope of a deterministic signal plus a narrow-band Gaussian noise process $\alpha(\tau)$ centered at some high frequency f_0

$$\alpha(\tau, \sigma) = \mathrm{Re}\left[a(\tau, \sigma)\exp(i2\pi f_0\tau)\right].$$

It is also meaningful to expand $a(\tau, \sigma)$ in a complete orthonormal Karhunen-Loeve series [11]

$$a(\tau, \sigma) = \sum_{i=0}^{\infty} a_i(\sigma)\phi_i(\tau) = \sum_{i=0}^{\infty} (s_i(\sigma) + n_i)\phi_i(\tau)$$

having the following properties:

1) The $\{\phi_i(\tau)\}$ are solutions to the integral equation

$$\lambda_i\phi_i(\mu) = \int_0^t K_n(\mu, v)\phi_i(v)\,dv$$

where

$$K_n(\mu, v) = E[n(\mu)n^*(v)]$$

in the covariance function of the noise and is a real function.

2) $a_i(\sigma) = \int_0^t a(\tau, \sigma)\phi_i^*(\tau)\,d\tau = (a, \phi_i) = (s, \phi_i) + (n, \phi_i)$.

3) The equality is in the sense of "limit-in-the-mean."

4) $(\phi_i, \phi_j) = \delta_{ij}$

5) The $a_i(\sigma)$ are independent Gaussian random variables, when conditioned on σ.

The generating function of this process N_t can then be written as [12]

$$M_{N_t}(s) = E\left[\exp\left(\alpha\int_0^t |a(t, \sigma)|^2 dt[e^\mu - 1]\right)\right]$$
$$= E\left[\exp\left(\alpha\sum_{i=0}^{\infty}|a_i(\sigma)|^2(e^\mu - 1)\right)\right]$$

which, using property (5) and [13] reduces to

$$M_{N_t}(s) = \prod_{i=0}^{\infty} E[\exp(\alpha|a_i(\sigma)|^2[e^\mu - 1])]$$
$$= \prod_{i=0}^{\infty}\frac{\exp(\alpha|s_i(\sigma)|^2(e^\mu - 1)/[1 - \alpha\lambda_i(e^\mu - 1)])}{1 - \alpha\lambda_i(e^\mu - 1)}. \quad (36)$$

At this point, the variable σ will be suppressed, although it must be considered as a conditioning variable when encountered in practice.

Notice that $M_{N_t}(s)$ is a product of similarly distributed functions. The inverse transform of the ith component is

$$P_{N_{t_i}}(k_i) = \frac{(\alpha\lambda_i)^{k_i}}{(1 + \alpha\lambda_i)^{1+k_i}}\exp\left(\frac{-\alpha|s_i|^2}{1 + \alpha\lambda_i}\right)L_{k_i}\left(\frac{-\alpha|s_i|^2}{\alpha\lambda_i(1 + \alpha\lambda_i)}\right) \quad (37)$$

where $L_x(y)$ is the Laguerre polynomial.

1) No Additive Noise: In the limit as $\lambda_i \to 0$ (37) approaches

$$\lim_{\lambda_i\to 0} P_{N_{t_i}}(k_i) = \frac{[\alpha|s_i|^2]^{k_i}}{k_i!}\exp(-\alpha|s_i|^2)$$

and

$$P_{N_t}(k) = \frac{\left[\alpha\sum_{i=0}^{\infty}|s_i|^2\right]^k}{k!}\exp\left(-\alpha\sum_{i=0}^{\infty}|s_i|^2\right)$$
$$= \frac{(\alpha E_S)^k}{k!}\exp(-\alpha E_S) \quad (38)$$

where $k = \sum_{i=0}^{\infty} k_i$ is the total count and $E_S = \sum_{i=0}^{\infty}|s_i|^2$ is the total signal energy in the $(0, t)$ interval. Thus the deterministic signal alone yields a Poisson distributed count. This, of course, could have been deduced immediately from (34). Notice, however, that when $|s_i|^2 = 0$

$$P_{N_{t_i}}(k_i) = \frac{(\alpha\lambda_i)^{k_i}}{(1 + \alpha\lambda_i)^{1+k_i}} \quad (39)$$

and each of the coordinate components is Bose-Einstein distributed [4].

In summary we see that: the signal alone can be considered to be Poisson distributed along each of its coordinate axes in Hilbert Space; Gaussian noise alone is Bose-Einstein distributed along a particular set of coordinate axes in Hilbert Space; when signal is added to the noise the resultant process is distributed according to (37) along each of the coordinate axes determined by the noise alone.

2) Band-Limited White Gaussian Noise: An important case occurs in communication theory when the signal and noise are passed through a filter before detection. We will consider the case where the process $a(\tau)$ is band limited by a rectangular filter with bandwidth $2B$. We will also assume that the noise was initially white, with spectral density N_0.

It has been shown [14] that when a process is band limited and then observed over a time interval $(0, t)$ the eigenfunctions are prolate spheroidal wavefunctions. It has also been shown that the first $(2Bt + 1)$ of these functions accurately approximate the original function. This appears valid for values of $2Bt$ as low as 3 and 5 [11]. Therefore, it is a good engineering approximation to assume that the eigenvalues associated with the first $(2Bt + 1)$ coordinates are each N_0 with the remaining ones being zero. The generating function $M_{N_t}(s)$ in (3) then becomes

$$M_{N_t}(s) \simeq \frac{\exp\left[\dfrac{\alpha(s, s)(e^\mu - 1)}{1 - \alpha N_0(e^\mu - 1)}\right]}{[1 - \alpha N_0(e^\mu - 1)]^{2Bt+1}} \quad (40)$$

with the corresponding probability density being

$$P_{N_t}(k) = \frac{(\alpha N_0)^k}{(1 + \alpha N_0)^{k+2Bt+1}}$$
$$\cdot\exp\left[\frac{-\alpha(s, s)}{1 + \alpha N_0}\right]L_k^{2Bt}\left[\frac{-\alpha(s, s)}{\alpha N_0(1 + \alpha N_0)}\right] \quad (41)$$

where $L_k^{2Bt}(x)$ is the Laguerre polynomial. We will now consider some limiting forms of (41).

3) No Signal: In the absence of signal, (41) reduces to

$$P_{N_t}(k) = \binom{2Bt + k}{k}\left(\frac{1}{1 + \alpha N_0}\right)^{2Bt+1}\left(\frac{\alpha N_0}{1 + \alpha N_0}\right)^k$$

which is a negative binomial distribution. There are two important limiting cases for this distribution.

a) Limit $2Bt \to 0$

$$P_{N_t}(k) = \frac{(\alpha N_0)^k}{(1 + \alpha N_0)^{k+1}}.$$

For $2Bt \ll 1$, there is only one significant eigenvalue, the average value. Since this occurs when $t \ll 1/2B$, it can clearly be related to the approximation

$$\int_0^t |a(\tau)|^2\,d\tau \simeq |a(0)|^2 t = \lambda_0 t \tag{42}$$

using the mean value theorem for integrals. This latter approximation is commonly used to obtain this result but lacks the insight as to the meaning or the range of validity.

b) Limit $2Bt$ large, $\alpha N_0 \ll 1$

$$P_{N_t}(k) = \frac{[\alpha 2BtN_0]^k}{k!} \exp(-\alpha 2BtN_0).$$

Notice that since $2N_0B$ is the noise power, $2Bt\alpha N_0$ is the total noise energy in the $(0, t)$ interval. If we write this as $\bar{I}t$, we have:

$$\int_0^t \alpha|a(t)|^2\,dt = \bar{I}t$$

where $\bar{I}$ is in fact the time-averaged noise power

$$\bar{I} = \frac{1}{t}\int_0^t \alpha|a(t)|^2\,dt.$$

Thus for large $2Bt$, there is a smoothing of the fluctuations in the noise process, and Poisson statistics prevail. The condition $\alpha N_0 \ll 1$ is a little difficult to interpret, except that it implies there be much less than one noise count per degree of freedom, which is easily obtained in practice. If one recognizes that a narrow optical filter has a bandwidth on the order of 1 Å at visible wavelengths, or about 100 GHz, it is clear that large $2Bt$ is the most common form of operation. $2Bt$ will be comparable in magnitude to the ratio of the optical filter and system bandwidths. Further, since almost all noise has a thermal origin,

$$\alpha N_0 = \frac{\eta}{\exp(h\nu/kT) - 1} \ll 1$$

is satisfied at optical frequencies. Actually, this is true assuming one mode of operation. However, for the purpose of this discussion we have considered a plane wave, or one spatial mode.

4) Signal Plus Noise: For this case, there are also two limiting conditions for (41).

a) Limit $2Bt \to 0$

$$P_{N_t}(k) = \frac{(\alpha N_0)^k}{(1 + \alpha N_0)^{1+k}} \exp\left[\frac{-\alpha(s, s)}{1 + \alpha N_0}\right] L_k\left(\frac{-\alpha(s, s)}{\alpha N_0(1 + \alpha N_0)}\right).$$

As in the case for no signal, the probability density reduces to that of an individual coordinate (37). Again this can be interpreted as the zero order eigenvalue or average value, as in (42).

b) Limit $2Bt$ large, $\alpha N_0 \gg 1$

$$P_{N_t}(k) = \frac{[\alpha\{2BN_0t + (s, s)\}]^k}{k!} \exp[-\alpha\{2BN_0t + (s, s)\}].$$

As might be expected from condition 3a) and (38), the limiting condition for large $2Bt$ and $\alpha N_0 \ll 1$ corresponds to a Poisson-distributed signal plus independent Poisson-distributed noise. Since this is the most common condition that one encounters in practice considerable effort has gone into exploring this approximation [17]–[20].

5) An Equivalent Eigenspace: Let us reexamine (37) and (41). Equation (41) is obtained as a $(2Bt+1)$-fold convolution of probability densities in (37), where all the λ_i are equal to N_0. This can be written as

$$P_{N_t}(k) = \bigotimes_{i=0}^{2Bt} \frac{(\alpha N_0)^{k_i}}{(1 + \alpha N_0)^{1+k_i}} \cdot \exp\left(\frac{-\alpha|s_i|^2}{1 + \alpha N_0}\right) L_{k_i}\left(\frac{-\alpha|s_i|^2}{\alpha N_0(1 + \alpha N_0)}\right) \tag{43}$$

where $\otimes_{i=0}^{2Bt}$ denotes a $(2Bt+1)$ fold convolution. Notice that the only way in which the signal enters is through the energy (s, s). Now

$$(s, s) = \int_0^t |s(\tau)|^2\,d\tau$$

and since the signal is band limited to $\pm B$, we can partition the $(0, t)$ interval into $(2Bt+1)$ equal Δt intervals where $(2Bt+1)\,\Delta T = t$. We can then closely approximate (s, s) as

$$(s, s) \simeq \sum_{j=0}^{2Bt} |s_j(j\Delta T)|^2 \Delta T, \quad j = 0, 1, 2, \cdots, 2Bt.$$

We can also write k as $k = \sum_{j=0}^{2Bt} k_j$ where k_j is the contribution of the jth interval to the total count k. Equation (41) can then be decomposed into a $(2Bt+1)$-fold convolution of the form

$$P_{N_t}(k) = \bigotimes_{j=0}^{2Bt} \frac{(\alpha N_0)^{k_j}}{(1 + \alpha N_0)^{k_j+1}} \cdot \exp\left(\frac{-\alpha|s_j|^2\Delta T}{1 + \alpha N_0}\right) L_{k_j}\left(\frac{-\alpha|s_j|^2\Delta T}{\alpha N_0(1 + \alpha N_0)}\right). \tag{44}$$

Notice that (44) is equivalent to (43) and would be identical if $|s_i|^2 = |s_j|^2\Delta T$ for all $i = j$. On the other hand, (44) is meaningful as representing a processable signal formed from independent samples as opposed to an abstract eigenspace.

For the particular case where the noise process is wide sense stationary and $2Bt$ is large (see, for example, [11]), one can approximate the eigenfunctions by harmonically related cissoids, and $|s_i|^2$ and N_0 represent the Fourier coefficients of the power density spectrum. Equations (43) and (44) then express the duality of signal processing and design in both time and frequency.

We can elaborate on this duality using the time-frequency representation first considered by Gabor [21] (see Fig. 1). The received process $a(\tau)$ considered, exists over the interval $(0, t)$, with frequency components primarily contained in the interval $(-B, +B)$. This is a Hilbert space of $(2Bt+1)$ dimensions which can be considered either as intervals of bandwidth $1/t$ in frequency or duration $1/2B$ in time. Hence we can observe the count k_j by looking in the time interval $(j/2B, j+1/2B)$ with a filter of bandwidth $2B$ or we can observe the count k_i by looking in the frequency

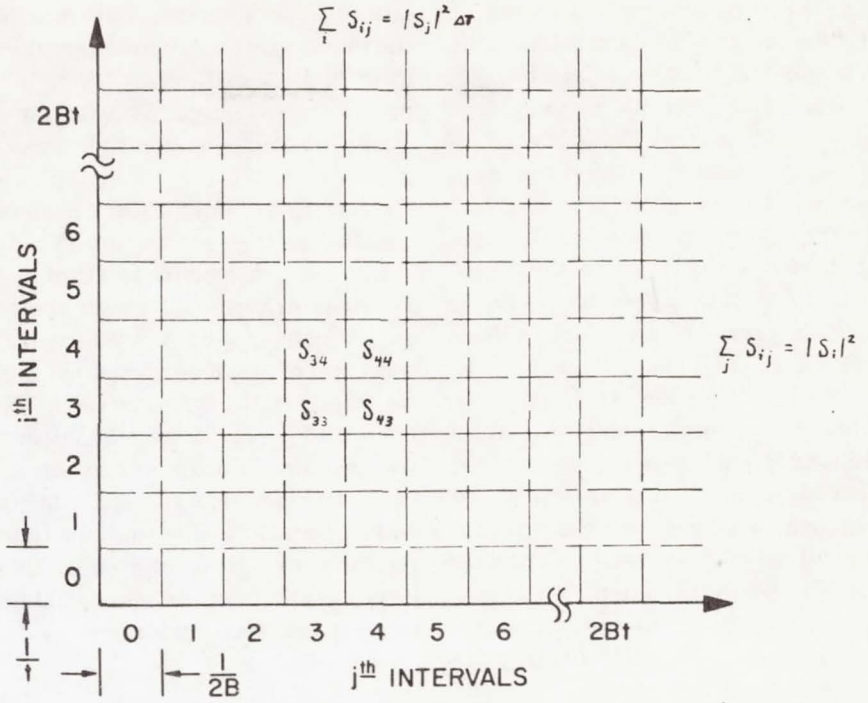


Fig. 1. Time-frequency representation in terms of $i-j$ intervals.

band $(-B+i/t, -B+(i+1)/t)$ for a time t. The first measurement is a sum of all the squares in the jth column, while the latter is a sum of all the squares in the ith row.

If the process is not wide sense stationary, we can still use Parseval's Theorem to write (s, s) as

$$(s, s) = \int_0^\infty P(f)\, df \simeq \sum_{l=0}^{2Bt} P(l\Delta f)\Delta f$$

and write a density similar to (44). K_t would be the total count in the band Δf in the interval $(0, t)$. However, one cannot assign the rigorous definition of power density spectrum to the noise and the noise coefficients.

We note, finally, that the most common statistical behavior encountered in practice yields $2B\Delta T \gg 1$. Hence condition 4b) applies to any measurement interval of length ΔT. Thus the observance of counts over many independent ΔT intervals is a sum of independent Poisson variables. This interpretation was first proposed by Reiffen and Sherman [17] on the heuristic basis, but can clearly be shown to have a solid foundation.

III. Shot Noise Processes

We have shown that a linear relation exists between the average power I of the radiation (over some finite aperture) and the rate of flow of photons n. Thus if n is a function of time, we can write

$$I(t) = h\nu n(t) \tag{45}$$

where h is Planck's constant and ν is the photon frequency. Thus the detector of optical radiation can be represented either as an instantaneous power detector or as an instantaneous rate detector. This relationship is generally explained by postulating that each incident particle independently releases an electron with probability η upon arrival at the photodetector surface, the electron in turn traveling to a cathode surface yielding a current "impulse" effect at the detector output. Thus the total output current $i(t)$ is due to the motion of a collection of electrons, proportional in number to the arriving particles. We can, therefore, write for the output current flow $i(t)$

$$i(t) = \sum_{m=1}^{N_t} h(t - t_m) \tag{46}$$

where $h(t)$ is the current impulse effect, t_m is the time of release of the mth photoelectron, and N_t is the number of such electrons occurring in the total time interval $(-\infty, t)$. The function $h(t)$ has area equal to the charge of an electron, while N_t is the counting statistic, discussed in Section II, of the photoelectron emissions. Note that if we neglect space-charge effects in the photodetector, the travel time of each released photoelectron is finite, which means that the function $h(t)$ must be time limited to some interval τ. That is, $h(t)=0$ for $t<0$ and $t>\tau$. In this case, N_t becomes the counting statistic over the finite interval $(t-\tau, t)$. Since τ is inversely related to the detector bandwidth, τ is relatively short $(10^{-10}-10^{-7}$ s), and can be considered a "delta function" with respect to most modulation waveforms. It perhaps should be pointed out that if $h(t)$ is assumed to be a flat rectangular function over $(0, \tau)$, then $i(t)=N_t-N_{t-\tau}$ and the detector output is precisely the counting process of the received optical radiation. If, instead, a nonrectangular impulse waveshape is to be accounted for, then one is forced into a closer examination of the processes described by (46). This class of processes can loosely be defined as "shot noise" processes (although the exact definition of the latter tends to vary in the literature).

As discussed in Section II, the parameter N_t is a random variable depending upon the intensity of the received field. Recall that if $n(t)$ is a deterministic function, N_t is a Poisson random variable, with mean value given by the integral of $\eta n(t)$ over $(t-\tau, t)$, and is a conditional Poisson random variable if the intensity $n(t)$ is a sample function of a continuous stochastic process. That is, given any intensity function of the ensemble, the counting process N_t is Poisson. With Poisson counting processes the resulting shot noise processes are referred to as Poisson shot noise (PSN). Some excellent discussions of PSN processes are given by Rice [11], Middleton [22], Papoulis [24], and Parzen [23]. In essence, first- and second-order statistics, such as probability densities, moments, power spectra, and correlation functions have been well developed. For the conditional PSN, the foregoing statistical characteristics can be formally attained by taking subsequent averages over the PSN results. For example, consider the power spectrum of the conditional PSN process in (46), where the intensity $n(t)$ is a sample function of an ensemble of positive random stationary process N defined over $(-\infty, \infty)$. We formally define the time averaged power density spectrum [25] of the shot noise process $i(t)$ by

$$\bar{S}_i(\omega) = \lim_{T\to\infty} \frac{1}{2T} E\left[|I_T(\omega)|^2\right] \tag{47}$$

where E is the expectation operator and $I_T(\omega)$ is the Fourier transform of $i(t)$ over $(-T, T)$. For the PSN processes, (47) can be readily determined as

$$\bar{S}_{\text{PSN}}(\omega) = \lim_{T\to\infty} \frac{1}{2T} \left[\bar{N}_T + |\Phi_T(\omega)|^2\right]|H(\omega)|^2 \tag{48}$$

where

$$\bar{N}_T = \int_{-T}^{T} n(t)\,dt \tag{49}$$

and $H(\omega)$ and $\Phi_T(\omega)$ are the Fourier transforms of $h(t)$ and $n(t)$, $-T<t<T$, respectively. The subsequent statistical average over N, and time average over T via the limiting operation, yield the power density spectrum for conditional PSN processes

$$\bar{S}_i(\omega) = |H(\omega)|^2\left[E(N) + S_N(\omega)\right] \tag{50}$$

where $S_N(\omega)$ is now the time averaged power density spectrum of the stochastic intensity $n(t)$. The foregoing results are significant, since it is valid for any counting statistic generated from conditional Poisson statistics and, therefore, includes those discussed in Section III. Note that the spectrum always takes the form of the intensity spectrum immersed in a background of "noise" of spectral shape $E(N)|H(\omega)|^2$. (For infinite bandwidth detectors, $H(\omega)\approx 1$ for all ω, and the aforementioned represents basically "white" noise.) This noise constitutes the shot noise of the detector, and is due to the discreteness of the photoelectron model. The intensity spectrum $S_N(\omega)$, in general, contains portions due to desired intensity modulation, portions due to background effects, and associated cross-spectral terms. These latter two components constitute the "fluctuation" noise of the photodetector output. Since the spectrum in (50) has the form of a "signal noise," there is a tendency to view the photodetected output as signal plus additive noise. The difficulty, of course, is that the signal and noise are not independent, and usual "signal plus noise" interpretations, familiar to communication engineers, often lead to false conclusions (e.g., see Section IV).

It is often instructive to examine the "instantaneous" or "short-term averaged" power spectrum of the detector output, which can be viewed basically as the conditional spectrum of (48) before the time averaging limit is taken. If we interpret the $2T$ interval to be the interval $(t-\tau, t)$, instead of $(-T, T)$, we see that the bracketed term in (48) will contain terms dependent on t. Furthermore, if we include the fact that the electron functions $h(t)$ have time widths τ much shorter than the time variations in $n(t)$, then the intensity $n_\tau(t)$ is approximately constant over $(t-\tau, t)$. Its power spectrum is then a delta function and the bracketed terms in (48) take the form

$$\left[k_\tau(t) + n^2(t)\left(\frac{\sin \omega\tau/2}{\omega\tau/2}\right)^2\right]|H(\omega)|^2.$$

That is, the instantaneous spectrum (power spectrum before the time average is taken) has the appearance of a background shot noise whose level varies with time, and whose average value varies according to the instantaneous value of $n(t)$. In this sense, the detector acts as an instantaneous "power" detector, which is the accepted classical definition of photodetectors. The true frequency content of the shot noise is not exhibited, however, until the time averaging is invoked.

The foregoing discussion raises an interesting query that cannot be answered from a spectral density point of view. If the shot noise process is to represent a true intensity detector, even when $n(t)$ is a stochastic process, then the statistical properties of the shot noise in (46) must be related to those of the intensity process $n(t)$. When the intensity is a deterministic time function, the relations between the shot noise and its intensity are well known. However, when the intensity is itself stochastic, the manner in which the statistics of the intensity and the conditional PSN are related is somewhat vague. For example, although the first-order probability density of $i(t)$ is difficult to write in closed form, its characteristic function is immediately available by making use of the known characteristic function of PSN [12], [23], [24]. Thus

$$\phi_{i_t}(\omega) = E_N\left[\exp \sum_{r=1}^{\infty} \frac{(j\omega)^r}{r!} \int_{t-\tau}^{t} n(\rho)h^r(t-\rho)\,d\rho\right] \tag{51}$$

where E_N is the average over the process N. One way to interpret (51) is to assume infinite bandwidth detectors, and factor the first term of the exponential summation. Thus

$$\phi_{i_t}(\omega) = E_N\{\exp(j\omega n(t))G[\omega, n(t)]\} \tag{52}$$

where the G function represents the remaining factors. The average of the first term alone is precisely the characteristic

function of the intensity process N at any time t. Thus the effect of the function G is to cause a departure of the first-order probability density of $i(t)$ from that of $n(t)$. The conditions under which the latter effect is negligible, and the shot noise process has approximately the first order density of N, have been studied by Karp and Gagliardi [26]. In this latter instance, we can say that the shot noise represents (statistically) the intensity process. This representation can be related to the "denseness" of the photon arrivals; i.e., the average number of photons per second. In fact, when the latter parameter is large, it can be shown that the bracketed term in (52) is approximately the characteristic function of a Gaussian random variable, with mean $n(t)$ and variance $n(t)$. This infers that the conditional (on N) probability density of $i(t)$ at any t approaches asymptotically a Gaussian density, which again may be loosely interpreted as an instantaneous "signal" $n(t)$, immersed in additive nonstationary Gaussian noise of variance $n(t)$.

The relation between shot noise and its intensity can be further investigated by consideration of the individual moments of the two processes. The moments of the process $i(t)$ can be obtained from its semi-invariants, which are, for PSN processes

$$\lambda_n(t) = \int_{t-\tau}^{t} h^n(t-\rho)n(\rho)\,d\rho. \tag{53}$$

The moments can then be obtained by the sequence of relations $E(i)=\lambda$, $E(i^2)=\lambda_2+\lambda_1^2$, $E(i^3)=\lambda_3+\lambda_1\lambda_2+\lambda_1^3$, etc. For conditional PSN processes, the λ_n are themselves random processes, and the moments of $i(t)$ depend upon the higher order moments of the process $n(t)$. However, if the intensities are continuous, or if the detector bandwidth is much larger than the bandwidth of the intensities, the rth moments are related by

$$E(i^r) = E(N^r) + D(r) \tag{54}$$

where $D(1)=0$ and $D(r)$, $r>2$, is function depending upon the higher order statistics of $n(t)$ and upon the function $h(t)$. This relationship was investigated in [26]. It was shown, for example, that if the function $h(t)$ was rectangular over $(0, \tau)$ the rth moment of $i(t)$ was approximately equal to the rth moment of the intensity process N if

$$\begin{bmatrix}\text{average number of}\\ \text{photon arrivals in}\\ \tau \text{ seconds}\end{bmatrix} \gg \frac{r(r-1)}{2}. \tag{55}$$

Equation (55) essentially states that the denseness of the shot events (i.e., the average number of $h(t)$ functions overlapping the time interval of one function) must be sufficiently large for moment representation. The right side of (55) serves as a rough rule of thumb for determining how large this denseness must be for approximate equality of the rth moment. It may be recalled [20] that for PSN processes (deterministic intensities) a condition of large number of shot occurrences is required before the PSN loses its discrete nature. Equation (55) can therefore be interpreted as the statistical equivalent of this statement; i.e., the condition under which the conditional PSN begins to take on the statistics of its intensity.

By using (54), it is also possible to relate the fluctuations in the detector output $i(t)$ to those of the intensity $n(t)$. Specifically, if we define the signal-to-noise ratio (SNR) of a positive process as the ratio of its mean value squared to the variance, then (54) leads to the fact

$$\text{SNR of } i(t) \leq \text{SNR of } n(t) \tag{56}$$

which implies that the percent fluctuations in the shot noise are always at least as great as those of the intensity itself. We make this point mainly because the foregoing definition of SNR is commonly used in assessing signal quality in communication system analysis.

It may be noted that the conditions for which the intensity is represented by a shot noise process are also useful in "building up" intensity models as shot noise. This type of shot noise modeling has been used for studying radiation scattering and perturbation effects [27], [28] in which the impulse functions $h(t)$ were reinterpreted as wave packets.

With the statistics of the conditional shot noise process identified (at least in first- and second-order statistics), the problem of optimal processing procedures at the photodetector output can now be properly formulated, and in some instances, solved. For example, the problem of optimal linear filtering of the process $i(t)$, so as to minimize the mean squared error from the desired intensity, was considered in [26]. For certain types of pulsed intensities, as in PCM communications optimal operations maximizing output signal to noise ratios have also been considered [29]. The application of estimation theory [30], tracking operations [31], and detection procedures [17], [18], [20] to the photodetector shot noise output has been under study, and appears to be a problem area of considerable interest from both a practical and theoretical point of view.

IV. Digital Communications and Optical Systems

The availability of an easily generated extremely narrow pulse in the optical region of the spectrum suggests a natural application to communication by digital methods. This notion, in turn, has fostered an increasing interest in the application of both classical detection theory and information to optical systems. Since the output of a photodetector is a sequence of electron counts, the detection problem is formally one of decisioning in the presence of generalized Poisson statistics. While early approaches to the problem basically were confined to pure Poisson counting [32], [33], more recent attention has included the generalized Laguerre counting processes in Section III, [13].

The formulation of the general M-ary detection problem involving counting statistics proceeds as follows. The transmitter sends a signal whose intensity is modulated with one of a set of M possible intensities, each T seconds long. The received signal is corrupted by background radiation, which we assume here is white Gaussian noise of level N_0 watts per hertz per unit area, and optical bandwidth B. The output of the photodetector at the receiver is then a

time varying process of electron counts, obeying a generalized Poisson distribution, as in Section III. The receiver observes the counting process over $(0, t)$ and decides which of the M possible intensities is being received. Since K binary digits can be uniquely encoded into $2^K = M$ possible intensity waveforms, a correct decision effectively decodes K data bits. The foregoing model can be cast into a discrete format by subdividing the interval T into ΔT-second intervals ($\Delta T \approx 1$/information bandwidth) and associating a signal energy component s_{ji} for the jth intensity and ith interval. (That is, s_{ji} is the total energy associated with the $2B\Delta T$ samples, or modes, of the Karhunen–Loeve expansion of the jth intensity during the ith ΔT interval.) Under a fixed energy constraint, we require $\sum_i s_{qi} = E$ for all q. The discrete problem then is to detect which of the possible intensity vectors $\boldsymbol{s}_q = \{s_{qi}\}$ is controlling the counting process by observing the sequence of independent counts $\boldsymbol{k} = \{k_i\}$, $i = 1, 2, \cdots, M(= T/\Delta T)$. Under a maximum likelihood detection criterion, and a priori equally likely signals, the optimal test is to form the likelihood functionals $\Lambda_q(\boldsymbol{k})$ and select $\boldsymbol{s}_q$ as being transmitted if no other $\Lambda_i(\boldsymbol{k})$ exceeds $\Lambda_q(\boldsymbol{k})$. If a likelihood draw occurs (more than one $\Lambda_q(\boldsymbol{k})$ is maximum) any randomized choice among the maxima can be used. From (37), the likelihood test is therefore equivalent to comparing:

$$\Lambda_q(\boldsymbol{k}) = \prod_{i=1}^{M} L_{k_i}^{2B\Delta T}\left(\frac{-s_{qi}}{N_0(1+\alpha N_0)}\right) \tag{57}$$

for all q, where s_{qi} is now a normalized signal intensity obeying the constraint $\sum s_{qi} = E \equiv N$. In typical operation, $2B\Delta T \gg 1$ (i.e., the optical bandwidth is much greater than the information bandwidth) and (1) is approximately

$$\Lambda_q(\boldsymbol{k}) \cong \prod_{i=1}^{M} \left[2B\Delta T + \frac{s_{qi}}{N_0(1+\alpha N_0)}\right]^{k_i} \Big/ k_i!. \tag{58}$$

After observing $\boldsymbol{k}$, examination of the set of $\{\Lambda_q\}$ for maxima is equivalent to the comparison of the set of functions $\sum k_i \log(1 + s_{qi}/K)$, where $K = 2BN_0\Delta T$ represents the noise energy per counting interval per unit area. (Recall it was previously shown in Section II that under the condition $2B\Delta T \gg 1$ the counts k_i are Poisson variates so that complete statistics of the foregoing test can be determined.)

An indication of the performance of a detection test is given by the divergence, or "expected distance between hypothesis." The divergence is formally defined as

$$D_{jq} = E_{\boldsymbol{k}}(\Lambda_{jq}|j) - E_{\boldsymbol{k}}(\Lambda_{jq}|q) \tag{59}$$

where $\Lambda_{jq} = \Lambda_j(\boldsymbol{k}) - \Lambda_q(\boldsymbol{k})$ and $E_{\boldsymbol{k}}(\Lambda|j)$ is the conditional average of Λ over $\boldsymbol{k}$ given the intensity $\boldsymbol{s}_j$. Abend [18] had shown that for Poisson counting, using the functions of (58) and $M = 2$ (binary detection), the divergence normalized by the variance of Λ is maximized by a "pulsed" type of intensity, in which the available signal energy is wholly concentrated in a single counting interval. That is, an intensity set defined by

$$\boldsymbol{s}_q = \{N\delta_{qi}\} \tag{60}$$

where δ_{qi} is the Kronecker delta function. Kailath [19] extended this result by showing that under a total energy constraint, other suitable forms of distance are maximized by similar pulsed intensities. Gagliardi and Karp [20] applied an average divergence criterion to the general M-ary Poisson detection problem and again showed the optimality of the intensity set of (60). In the latter reference, the intensity set that maximized the probability of correctly detecting the true intensity, rather than maximizing divergence, was also considered, and shown to correspond to the pulsed set in two special cases, 1) $M = 2$ with symmetric intensity sets and 2) any M and low intensity-to-noise-energy ratio. However, the determination of global optimal intensity sets in the pure Poisson case, based upon detection probability still remains a difficult task. It has been conjectured by many that the pulsed set of (60) is, in fact, a global optimal set, but to the authors' knowledge a rigorous proof has not been shown. The optimality of the pulsed set, even under this special criterion, is significant, since it indicates the importance of intensity waveshape in digital system design. This, of course, is partly due to the general advantage of orthogonal signals in detectability, a property afforded by the disjointness of the pulsed set in (60). The use of signals placed in adjacent time slots is in essence a pulse position modulated system in which each position corresponds to a digital word. The dual of such a system (a frequency keyed system), which also retains the orthogonality property, can similarly be generated by redefining the expansion functions of the received field [34].

It should be pointed out that if the condition $2B\Delta T \gg 1$ is not valid, care must be used in accepting the pulsed set of (60) as an optimal intensity set. In particular, the Poisson assumption and the use of (58) is violated. For the case of $2B\Delta T \ll 1$, the divergence in (59) must be obtained by averaging terms as in (57) over the Laguerre densities. If this averaging is carried out, (59) takes the form

$$D_{qj} \cong C\left\{\prod_{i=1}^{M} I_0\left(\frac{s_{qi}}{N_0}\right) + \prod_{i=1}^{M} I_0\left(\frac{s_{ji}}{N_0}\right) - 2\prod_{i=1}^{M} I_0\left(\frac{s_{qi}s_{ji}}{N_0}\right)\right\} \tag{61}$$

where $I_0(x)$ is the imaginary Bessel function of zero order and C contains terms common to all q and j. Now it is no longer immediately evident that the pulsed set of (60) maximized D_{qj}. The last term, however, is minimized if either $s_{qi} = 0$ or $s_{ji} = 0$ for all i, which suggests a disjoint intensity set, but it is not evident that the first terms are maximized under the same condition. The difficulties of this problem are quite reminiscent of similar difficulties in attempting to find optimal signal sets in noncoherent additive Gaussian noise channels.

When the pulsed set of (60) is used and the general Laguerre counting is assumed, the analysis procedures are similar to the Poisson case. It is easy to show the monotonicity of Laguerre functions with respect to their indices. It then follows from (57) that $\Lambda_q \gtrless \Lambda_i$ if $L_{k_q}^{a}(N/N_0)$

$\gtrless L_{k_i}^{a}(N/N_0)$ which, in turn, is true if $k_q \gtrless k_i$. Hence the maximum likelihood test need only count over each interval, selecting the signal corresponding to the interval with the largest count.

A. Error Probabilities with Pulsed Intensity Sets and Poisson Counting

The performance of the pulsed intensity set in M-ary detection can be evaluated by considering the error probability when Poisson counting statistics are assumed. This can be obtained by noting that for the pulsed intensity set of (60) the log of the likelihood functions for each k_i constitutes a set of independent Poisson random variables. The variable for k_q has intensity $(N+2B\alpha N_0\Delta T)$ if the qth intensity was sent, and has intensity $K=2B\alpha N_0\Delta T$ otherwise. Recall that if the qth intensity is sent a correct decision will be made with probability $1/(r+1)$ of the log likelihood equals r others and exceeds the remaining $M-(r+1)$. Therefore, upon considering all possibilities, the error probability can be derived as [20]

$$\dot{P}_E(E, K, M) = 1 - \frac{e^{-(N+MK)}}{M} - \sum_{x=1}^{\infty}\left[\frac{(N+K)^x e^{-(N+K)}}{x!}\right] \cdot \left[\sum_{t=0}^{x-1}\frac{K^t e^{-K}}{t!}\right]^{M-1} \cdot \left[\frac{(1+a)^M - 1}{aM}\right] \quad (62)$$

where:

$$a = \left[\frac{K^x}{x!\sum_{t=0}^{x-1}\frac{K^t}{t!}}\right].$$

The function $P_E(N, K, M)$ has been plotted by Pratt [32] for $M=2$, and recently a digital computation has been generated [23] for a complete plot of the function. An exemplary plot is shown in Fig. 2 in which $P_E(N, 3, M)$ is plotted for various M as a function of N. It is important to note that P_E depends on both the normalized signal energy N and the normalized noise energy K in the counting interval, and not simply on their ratio. This fact is emphasized in Fig. 3, in which $P_E(N, K, 2)$ is plotted as a function of K for 2 fixed ratios N/K. This dependence on both signal and noise energies distinguishes the Poisson detection problem from the analogous coherent Gaussian channel problem. Note that the interfering noise energy K depends only upon the background energy in the interval ΔT, which is the width of the transmitted intensity pulse. The prime advantage of Poisson systems is precisely their ability to remove the effect of background noise by making ΔT small, and has been emphasized in previous reportings [36], [37].

The actual dependence of P_E on the parameter ΔT has been considered [38], and the improvement in error probability with decreasing ΔT has been demonstrated. The improvement, of course, is made at the expense of information bandwidth and peak power, both inversely proportional to ΔT. Surprisingly, the improvement is quite small at low values of N, and the increase in bandwidth may not be worth the decrease obtained in error probability. The effect

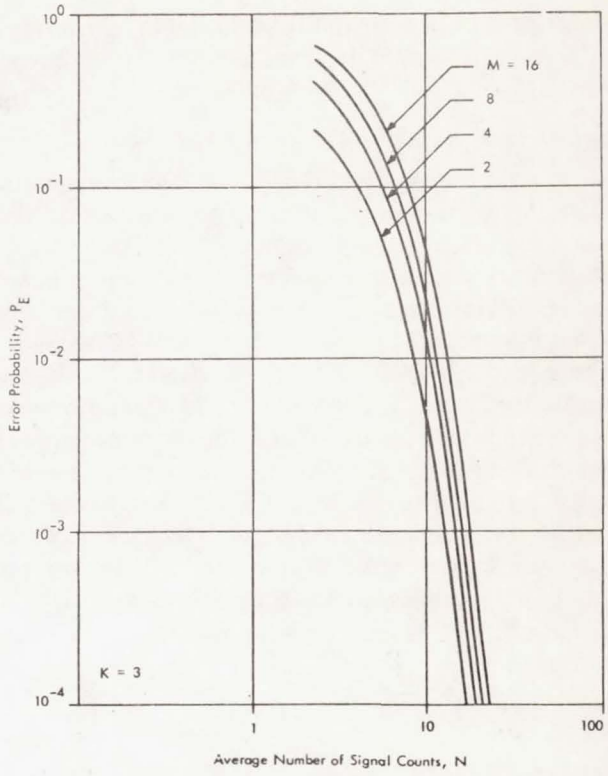


Fig. 2. Error probabilities for M-ary signaling.

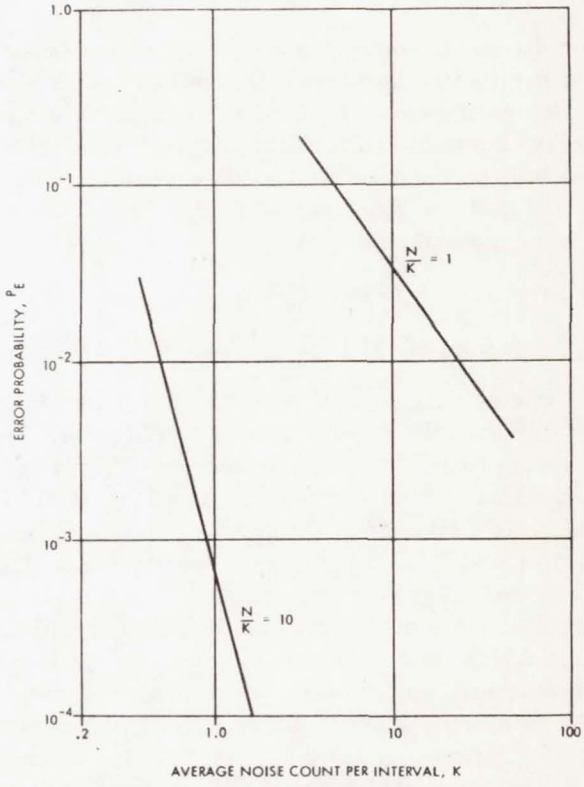


Fig. 3. Error dependence on signal and noise energies.

on error probability of additive extraneous thermal noise in the decisioning system and statistical characteristics of photomultipliers has also been considered [38].

For Laguerre counts, (62) must be rederived using the Laguerre densities discussed in Section III. Recently, general

bounds or the error probability in this latter case, using the orthogonal (disjoint) signal intensity sets, have been reported [34].

B. Information Rate of a Poisson PPM System

We have so far analyzed only one aspect of system performance, i.e., error probabilities. The actual information rate that the link achieves is another important design consideration. As stated, the transmitter sends optical energy in one of M time intervals, which is ΔT seconds wide, thereby transmitting one of M possible signals in $M\Delta T$ seconds, or at a rate $\log_2 M/M\Delta T$ bit/s. The receiver correctly determines the true signal with probability $1-P_E$ and is in error with probability P_E. Because of symmetry, the erroneous signal may be equally likely interpreted as any of the $M-1$ incorrect signals. Thus the overall channel may be depicted as an M-ary symmetric channel, in which each of the M possible transmitter signals is converted to itself with probability $1-P_E$ and converted to each other signal with probability $P_E/(M-1)$. The information rate for such a channel is known to be

$$H=\frac{\log_2 M+P_E\log_2\left[P_E/(M-1)\right]+(1-P_E)\log_2(1-P_E)}{M\Delta T}. \quad (63)$$

For convenience we shall denote this as

$$H = C(N, K, M)/M\Delta T \quad (64)$$

to emphasize the dependence of the numerator on the stated parameters. By using (63) and the families of error probability curves as in Fig. 2, the rate H can be evaluated by straightforward substitution. Although specific curves for such a computation are not shown here, it suffices to note that if N and K are such that $P_E<10^{-1}$, then (63) is, to a good approximation

$$\begin{aligned} H &\approx (1-P_E)\left[(\log_2 M)/M\Delta T\right] \\ &= (\log_2 M)/M\Delta T - P_E\left[(\log_2 M)/M\Delta T\right]. \end{aligned} \quad (65)$$

If we interpret the rate H as the source rate minus the equivocation of the channel, then the PPM optical system behaves approximately as if a source rate of $\log M/M\Delta T$ is passed into a channel of equivocation $P_E \log M/M\Delta T$. As noted in (62), even if $K\to 0$ (no background interference), $P_E\to\exp(-N)/2$, so that the equivocation is not due entirely to the background noise.

The use of (63) and the previous equations are helpful in determining the rate, given operating parameters. However, the converse design problem, which is to determine particular parameter values that achieve a desired rate, is not so straightforward. This is due to the fact that the rate is a somewhat complicated function of the parameters. We shall consider here two aspects of this design problem that have practical application under certain operating conditions. First, the word period $T=M\Delta T$ is held fixed while the information bandwidth $1/\Delta T$ is allowed to vary, and second, the bandwidth is held fixed while the word period is allowed to vary. In both cases, we are interested in the relationship between the rate H and the transmitter parameters N and M, assuming that the noise power is held fixed.

C. Fixed Word Period

We assume here that ΔT is allowed to vary with M so as to maintain $T=M\Delta T$ constant. Thus the system "squeezes" more signals into the T-second period as M increases. The resulting rate is then

$$H = C(N, K_T/M, M)T \quad (66)$$

where K_T is the noise energy in T. Thus the rate depends only upon the numerator of (63). With N fixed, increasing M increases the source rate, but the error probability also increases and eventually reaches an asymptotic value of

$$P_E=\left(1+\frac{N\left[K_T-1+\exp(-K_T)\right]}{K_T}\right)\exp(-N)$$

for large M. The resulting system rate increases, to within a constant of the entropy of the alphabet, $\log_2 M/T$. Therefore, it is clear that if the bandwidth is expendable, one will always increase the system rate for large M by increasing M. In a practical system, this implies that one should operate with as wide a bandwidth as possible to fully exploit the capability of the PPM system. We are, therefore, led naturally to consider the design of a system for an arbitrary rate H, when the full bandwidth $(1/\Delta T)$ of the system is limited.

D. Fixed Bandwidth

In this case, ΔT is held constant (thereby fixing the noise energy K in ΔT) so that both the numerator and denominator in (63) depend upon M, and the rate degrades quickly as M increases due to the $\log M/M$ dependence. A given rate, e.g., H_0, may be obtained by many different combinations of N and M. Analytically, these equivalent operating points may be obtained graphically by noting that they are the values for which the numerator $C(N, T, M)$, considered as a function of M, intersects the straight line $H_0\Delta TM$. By plotting these functions for various N, their intersection will identify (N, M) pairs which achieve the rate H_0. One may then decide on a particular operating point by invoking suitable design criteria. For example, one may select the smallest M from among the candidate pairs, which then minimizes the word period $T=M\Delta T$. Alternatively, one may choose to minimize the average transmitter power per information bit, which is proportional to N/C. In the latter case, therefore, one would select the operating pair (N, M) for which N/C is minimal. The latter parameter is recognized as the β-efficiency parameter (energy per data bit) of a communication system [39]. If the value of N/C, corresponding to the optimal (N, M) pair, is tabulated, the results can be compared to previously derived performance based upon the same parameter. This type of comparison was considered [40] and it was shown that the PPM system outperformed an optical heterodyne system for sufficiently large M, approaching in fact the minimum β generated by the Gordon bound for quantum systems. This type of result further emphasizes the importance of expending system bandwidth (increasing M also implies increasing information bandwidth) to improve overall performance. The effect of Laguerre statistics (when the information band-

width approaches the optical bandwidth) and the effect of additive noise can be accounted for by modifying these Poisson results [40].

The extension of the discrete model for optical detection (which was assumed almost entirely in the aforementioned references) to the continuous model has received little attention. In usual procedures, the continuous case is generated from the discrete by taking limits of infinitely small intervals. Although this procedure can be properly structured to generate the continuous version of the counting process, the continuous process representing the photodetector output must be viewed entirely as a shot noise process (see Section III). Unfortunately, such processes have first-order densities that are expressible only through transforms of their characteristic function. Hence the building up of a general detection model based upon shot noise rather than discrete processes would be severely hampered by the inability to express observable statistics. It would appear, however, that shot noise detectability cannot continue to be avoided when consideration is given to operation with information bandwidths on the order of optical detector bandwidths. This aspect of detection deserves more attention in future research studies.

V. Analog Communications

The major portion of work in the area of analog communications for optical systems has centered on first- and second-moment theory, spectral analysis, and signal-to-noise ratios. We have already discussed spectral analysis for shot noise processes with emphasis on signal representation. For the remainder of the paper we will concentrate on trying to bring together some of these ideas in a unified way, leaning heavily on physical motivation.

Before turning to the analyses required it is very instructive to reconsider the behavior of a photodetector from a phenomenological point of view. As we have already seen an important parameter in a photodetector is the time ΔT over which the intensity fluctuations remain relatively constant. This is related to the bandwidth B of the optical signal by $\Delta T \leq 1/2B$. When an electron is released from the detecting surface and flows through the ensuing circuitry, there is always the fixed electron charge e. This fixes the area of the resulting current pulse. Hence higher energy electrons will flow faster, the current pulses will be narrower in time resulting in an increased frequency response of the detector.

Generally, one thinks of counting circuitry as literally counting each of these events. On the other hand, one can also consider the following viewpoint: suppose we "match" the detector response B_d to the incident radiation bandwidth $\Delta T \simeq 1/2B = 1/2B_d$. Then each current pulse created will be approximately ΔT seconds wide. Hence at any time t, the effects of all pulses from the previous ΔT seconds will still be present. Therefore, if k_i electrons flow in the interval $(t_i - \Delta T, t_i)$, then at the time t_i the value of the current can be approximated by $k_i(e/\Delta T)$, or since $k_i \Delta T \simeq \alpha \tilde{I}(t_i)$, $i(t_i) \simeq \alpha e \tilde{I}(t_i)$, which was shown earlier. If the response of the detector were square pulses, this description would be exact. On the other hand, the distortions occurring due to end effects are the normal effects of filtering. The so-called shot noise represents the fact that K_i is an integer, making $\tilde{I}(t_i)$ take on discrete values, whereas the true $I(t)$ would be continuous.

The previous argument was intended to justify consideration of the $(2Bt+1)$ Nyquist samples for analog processes also. It was shown in (55) that these samples can also be considered statistically independent.

1) *Maximizing Signal-to-Noise Ratio for Direct Detection:* For maximum likelihood detection, the optimum form of processing consisted of weighting the counts on each of the $(2Bt+1)$ intervals. We will, therefore, consider the form of processing where each k_j is weighted by the number β_j. The processed signal then becomes v, where

$$v = \sum_{j=0}^{2Bt} \beta_j K_j. \tag{67}$$

As a criterion for signal processing, we will use the signal-to-noise ratio defined as

$$\frac{S}{N} = \frac{E^2[v]|_{N_0}}{\text{var}[v]} = 0. \tag{68}$$

Thus the mean of v in the absence of noise can be obtained from (37) and is

$$E[v]|_{N_0=0} = \alpha \sum_{j=0}^{2Bt} \beta_j |s_j(j\Delta T)|^2 \Delta T \tag{69}$$

with the variance being

$$\text{var}[v] = \alpha \sum_{j=0}^{2Bt} \beta_j^2 \{(|s_j(j\Delta T)|^2 + N_0') + \alpha(N_0'^2 + 2N_0'|s_j(j\Delta T)|^2)\Delta T\}\Delta T \tag{70}$$

and $N_0' = N_0/\Delta T$.

Thus the signal-to-noise ratio becomes

$$\frac{S}{N} = \frac{\left\{\alpha \sum_{j=0}^{2Bt} \beta_j |s_j(j\Delta T)|^2 \Delta T\right\}^2}{\alpha \sum_{j=0}^{2Bt} \beta_j^2 \{(|s_j(j\Delta T)|^2 + N_0') + \alpha(N_0'^2 + 2N_0'|s_j(j\Delta T)|^2)\Delta T\}\Delta T} \tag{71}$$

which can be bounded by using the Schwarz inequality. Hence

$$\frac{S}{N} \leq \sum_{j=0}^{2Bt} \frac{\alpha\{|s_j(j\Delta T)|^2\}^2 \Delta T}{|s_j(j\Delta T)|^2 + N_0' + \alpha(N_0'^2 + 2N_0'|s_j(j\Delta T)|^2)\Delta T} \tag{72}$$

with the equality holding when

$$\beta_j = \frac{|s_j(j\Delta T)|^2}{|s_j(j\Delta T)|^2 + N_0' + \alpha(N_0'^2 + 2N_0'|s_j(j\Delta T)|^2\Delta T}. \quad (73)$$

Notice that in the absence of noise $N_0 = 0$

$$\left(\frac{S}{N}\right) \leq \alpha \sum_{j=0}^{2Bt} |s_j(j\Delta T)|^2\Delta T = \alpha E_s = \frac{\eta E_s}{h\nu}. \quad (74)$$

This, however, is the average number of photoelectron counts in the $(0, t)$ interval and is generally referred to as the quantum-limited signal-to-noise ratio.

Let us now rewrite the right-hand side of (72) as

$$\left(\frac{S}{N}\right) \leq \sum_{j=0}^{2Bt} \frac{\alpha|s_j(j\Delta T)|^2\Delta T}{1 + \alpha N_0 + [1/\alpha|s_j(\Delta T)|^2\Delta T]\{\alpha N_0' + \alpha^2 N_0^2\}} \leq \alpha E_s. \quad (75)$$

Recall now that αN_0 is the number of noise counts per ΔT interval and for thermal noise sources is much less than one. In addition, $\alpha|s_j(j\Delta T)|^2\Delta T$ is the average number of signal counts in the jth ΔT interval. Suppose, therefore, that we construct a signal

$$|s_j(\Delta T)|^2 = \frac{E_s}{\Delta T}, \quad \text{for one value of } j$$
$$= 0, \quad \text{for all other values of } j.$$

Then clearly

$$\sum_{j=0}^{2Bt} |s_j(\Delta T)|^2\Delta T = E_s$$

is not violated, and in addition

$$\left(\frac{S}{N}\right) = \alpha E_s/1 + \alpha N_0 + \left[\frac{\alpha N_0 + (\alpha N_0)^2}{\alpha E_s}\right] \simeq \alpha E_s \quad (76)$$

for all values of $\alpha E_s > \alpha N_0$. Thus low duty-cycle operation is preferable when maximizing the signal-to-noise ratio of detected radiation in the absence of detector noise.

The addition of independent thermal noise with temperature T at the detector output changes the variance in (70). After some manipulation to take into account the electron charge e, the bandwidth, and the load R, the signal-to-noise ratio in (76) can be written as

$$\frac{S}{N} = \alpha E_s/1 + \alpha N_0 + \left[\frac{\alpha N_0 + (\alpha N_0)^2}{\alpha E_s}\right] + \frac{kT}{e^2\alpha E_s R B_d}. \quad (77)$$

The quantity $kT/e^2\alpha E_s R B_d$ is, in general, much greater than one. Therefore, except under extreme conditions of temperature, impedance, bandwidth, and signal level, a normal detector will be "thermal noise limited" in operation and S/N will be much less than αE_s.

We have been considering the case where each sampling interval represented one mode. If, in fact, each interval contained L modes, then clearly we need only replace αN_0 by $L\alpha N_0$ everywhere.

2) *Direct Detection with Photomultiplication:* We have just shown that most detectors are inherently thermal noise limited except under extreme temperature conditions. This was true because the current that was released by the surface immediately encountered a thermal environment. There are devices, presently limited to the visible region of the spectrum, which impart a preamplification to the photocurrent before the thermal environment is met. The most common device, a photomultiplier tube, consists of a cascade of stages through which each emitted electron passes and is amplified many thousands of times. When the effect of an electron emitted at the cathode reaches the anode, it appears as an actual current pulse well above the anode thermal environment. It is, therefore, possible to view the effects of individual electrons. These devices are commonly referred to as "photon counters."

To first order, one can account for this amplification A by assuming an electron charge equal to Ae. Then we can see from (77) that the term which previously made the device thermal noise limited becomes

$$\frac{kT}{A^2e^2\alpha E_s R B_d}.$$

Thus if the gain of the device is such that the inequality

$$A > \sqrt{\frac{kT}{e^2(\alpha E_s)RB_d}} \approx 3 \times 10^6 \sqrt{\frac{T}{(\alpha E_s)B_d}}$$

is satisfied, it is again shot-noise limited. In practice, the gain is a random variable and an "excess noise" appears because of the finite variance of A. This, however, only causes changes on the order of 20 percent or about 1 dB, and for the purposes of this discussion can be ignored.

3) *Heterodyne Detection:* If the electric field of a local oscillator is aligned coincident with the received signal over the detector surface, then one can directly add the two electric fields. Thus if we designate the signal by $E_1 \exp(j\omega_1 t + \phi(t))$ and the local oscillator by $E_{LO} \exp(j\omega_2 t)$ then

$$s_j(j\Delta T) = E_1 \exp[j\omega_1(j\Delta T) + \phi(j\Delta T)] + E_{LO} \exp[i\omega_2(j\Delta T)]$$

and

$$|s_j(j\Delta T)|^2 = |E_1|^2 + |E_{LO}|^2 + 2|E_1||E_{LO}| \cdot \cos\{(\omega_1 - \omega_2)j\Delta T + \phi(j\Delta T)\}. \quad (78)$$

If the local oscillator is made large, then it can be shown that, under these conditions, the density in (41) approaches a Gaussian density with a mean value of $2\alpha|E_1||E_{LO}|$ cos $\{(\omega_1 - \omega_2)j\Delta T + \phi(j\Delta T)\}$ (excluding the dc component) and variance $\alpha|E_{LO}|$ multiplied by the bandwidth considered. Then if the bandwidth of the signal in (78) is $2W$ and the bandpass of the detector is greater than $(\omega_1 - \omega_2) + W$, one can pass the detected signal through a bandpass filter centered at $(\omega_1 - \omega_2)$ with bandpass $2W$ and recreate the signal $2\alpha E_1 E_{LO}$ cos $\{(\omega_1 - \omega_2)j\Delta T + \phi(j\Delta T)\}$. The resulting carrier signal-to-noise ratio will be

$$\left(\frac{S}{N}\right)_{\text{het}} = \frac{\frac{1}{2}(2\alpha|E_1||E_{LO}|)^2}{\alpha|E_{LO}|2W} = \frac{\alpha|E_1|^2}{W} = \frac{\eta|E_1|^2}{h\nu W}$$

which can again be recognized as the quantum limited condition.

4) Power Spectrum Analysis: In Section III it was shown that the time-averaged power density spectrum of the current could be written as

$$\bar{S}_i(\omega) = |H(\omega)|^2[E(N) + S_N(\omega)].$$

Since $S_N(\omega)$ is the spectrum of a nonnegative definite function (the normalized power), it can be written in terms of a dc and an ac component. The ac component is $S_{AC}(\omega)$ where

$$S_{AC}(\omega) = (\eta\bar{n}e)^2\Phi_M(\omega)$$

and $n(t)$ has been normalized to

$$n(t) = \bar{n}(1 + m(t)); \qquad m(t) \geq -1$$

with

$$\int_{-T}^{T} m(t)\,dt = 0$$

and $\Phi_M(\omega)$ the time-average power density spectrum of $m(t)$. Notice that the modulation index is included in $m(t)$. For an unmodulated source, such as noise, $m(t) \equiv 0$, and only the shot noise term and the dc remain. Thus if we have a signal plus additive noise impinging on the detector, where the average noise rate is designated $\bar{n}_n$, the power density spectrum minus the dc terms can be written as

$$S_T(\omega) = e^2|H(\omega)|^2[\eta(\bar{n}_n + \bar{n}) + (\eta\bar{n})^2\Phi_M(\omega)] + \frac{2kT}{R}$$

where we have also included the thermal noise contribution. If we define the signal-to-noise ratio as the ratio of the total signal power

$$\frac{(e\eta\bar{n})^2}{2\pi}\int_{2W'}\Phi_M(\omega)\,d\omega$$

over the bandwidth of the signal, divided by the total nonsignal power over the same bandwidth:

$$\frac{1}{2\pi}\int_{2W'}\left[e^2|H(\omega)|^2\eta(\bar{n} + \bar{n}_n) + \frac{2kT}{R}\right]d\omega$$

then, assuming that $|H(\omega)|^2$ is "flat" over the $2W'$ region of interest

$$\frac{S}{N} = \frac{(\eta\bar{n}e)^2\left\{\frac{1}{2\pi}\int\Phi_M(\omega)\,d\omega\right\}}{\left[e^2\eta(\bar{n} + \bar{n}_n) + \frac{2kT}{R}\right]2W} \leq \frac{(\eta\bar{n})}{\left[1 + \frac{\bar{n}_n}{\bar{n}} + \frac{2kT}{e^2R\eta\bar{n}}\right]2W} \tag{79}$$

where W is now the cyclic frequency. Notice again that if e is replaced by Ae and the shot noise term $2\eta A^2e^2\bar{n}W = 2\eta AeI_{DC}W > 4kTW$, the device will be again shot-noise limited. The term

$$\frac{\eta\bar{n}}{2W} = \frac{\eta P}{2h\nu W}$$

can again be recognized as being related to the quantum limited condition.

VI. Summary Remarks

We have tried to present in this paper a review of the basic concepts in optical communications viewed strictly from a classical point of view, in the absence of any channel effects. In this vein, we have viewed the received signal as an electromagnetic field and described its interaction with a photodetector. We then described some of the fundamental properties of the resulting current flow as seen by the communications engineer.

The treatment in this paper is not complete, since the study of this problem has not finished. Consequently, some portions have been given more emphasis than others, while some have been omitted entirely. For example, in the literature the topic of continuous estimation for shot noise processes has barely been touched [13]. The same is true for synchronization in a shot noise environment [31], although this will be fundamental to any sophisticated optical communications system.

What has been attempted, rather, was a presentation which answered the questions concerning the physical modelling of the system and a reduction to the terms most useful for analysis. Where such analysis had reached a level of conveying a reasonably complete understanding of an aspect of the problem, it was also presented. It is hoped that this paper is thorough enough to motivate additional research in this area.

References

[1] R. J. Glauber, "The quantum theory of optical coherence," *Phys. Rev.*, vol. 130, pp. 2529–2539, June 1963.

[2] ——, "Coherent and incoherent states of the radiation field," *Phys. Rev.*, vol. 131, September 1963.

[3] ——, "Optical coherence and photon statistics," in *Quantum Optics and Electronics*, C. de Witt, *et al.*, Eds. New York: Gordon and Breach, 1965, pp. 65–185.

[4] L. Mandel and E. Wolf, "Coherence properties of optical fields," *Rev. Mod. Phys.*, vol. 37, pp. 231–287, April 1965.

[5] J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics*. New York: W. A. Benjamin, Inc., 1968.

[6] G. J. Troup, *Optical Coherence Theory*. London, England: Methuen, 1967.

[7] W. H. Louisell, *Radiation and Noise in Quantum Electronics*. New York: McGraw-Hill, 1964.

[8] L. Mandel, E. C. G. Sudarshan, and E. Wolf, *Proc. Phys. Soc.*, vol. 84, 1964.

[9] E. Fermi, *Rev. Mod. Phys.*, vol. 4, 1932.

[10] V. Fano, *Amer. J. Phys.*, vol. 29, 1961.

[11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.

[12] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, 1944.

[13] S. Karp and J. R. Clark, "Photon counting: a problem in classical noise theory," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 672–680, November 1970.

[14] D. Slepian and H. O. Pollak, "Prolate spheroidal wave function, Fourier analysis, and uncertainty, I," *Bell Syst. Tech. J.*, vol. 40, pp. 43–64, 1961.

[15] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, II," *Bell Syst. Tech. J.*, vol. 40, pp. 65–84, 1961.

[16] ——, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, III," *Bell Syst. Tech. J.*, vol. 41, pp. 1295–1336, 1962.

[17] B. Reiffen and H. Sherman, "An optimum demodulator for Poisson processes: photon source detectors," *Proc. IEEE*, vol. 51, pp. 1316–1320, October 1963.

[18] K. Abend, "Optimum photon detection," *IEEE Trans. Inform. Theory* (Correspondence), vol. 12, pp. 64–65, January 1966.

[19] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. COM-15, pp. 52–60, February 1967.
[20] R. M. Gagliardi and S. Karp, "*M*-ary Poisson detection and optical communications," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 208–216, April 1969.
[21] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.* (Tokyo), vol. 93(III), pp. 429–457, November 1946.
[22] D. Middleton, *Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960.
[23] E. Parzen, *Stochastic Processes*. San Francisco, Calif.: Holden-Day, Inc., 1962.
[24] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1965.
[25] *Ibid.*, ch. 10.
[26] S. Karp and R. M. Gagliardi, "On the representation of a continuous stochastic intensity by Poisson shot noise," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 142–147, March 1970.
[27] S. Karp, R. M. Gagliardi, and I. S. Reed, "Radiation models using discrete radiator ensembles," *Proc. IEEE*, vol. 56, pp. 1704–1711, October 1968.
[28] D. Middleton, "A statistical theory of reverberation and similar first-order scattered fields—Part I: Waveforms and the general process; Part II: Moments, spectra, and special distributions," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 372–414, July 1967.
[29] R. W. Chang, "Photon detection for an optical pulse code modulation system," *IEEE Trans. Inform. Theory* (Correspondence), vol. IT-15, pp. 725–728, November 1969.
[30] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 31–37, January 1969.
[31] R. M. Gagliardi, "The study of synchronization techniques for optical communication systems," University of Southern California, Elec. Eng. Dept., 2nd, 3rd, and 4th Quarterly Reps., April, August, and December 1969.
[32] W. K. Pratt, "Binary detection in an optical polarization modulation communication channel," *IEEE Trans. Commun. Technol.* (Concise Papers), vol. COM-14, pp. 664–665, October 1966.
[33] M. Ross, *Laser Receivers*. New York: Wiley, 1966.
[34] J. W. S. Liu, "Reliability of quantum-mechanical communication systems," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 319–329, May 1970.
[35] S. Karp *et al.*, "Error probabilities for Poisson detection," NASA Tech. Note TN-D-4721, October 1968.
[36] M. Ross, "Pulse interval modulation laser communications," presented at the Eastcon Conv., Washington, D. C., October 1967.
[37] S. Karp and R. M. Gagliardi, "A low duty cycle optical communication system," presented at the Eastcon Conv., Washington, D. C., October 1967.
[38] ——, "The design of a pulse-position modulated optical communication system," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 670–676, December 1969.
[39] R. W. Sanders, "Communication efficiency comparison of several communication systems," *Proc. IRE*, vol. 48, pp. 575–588, April 1960.
[40] S. Karp, "Communication efficiency of quantum systems," NASA Tech. Rep., TR-R-R-320, September 1969.

The Design of a Pulse-Position Modulated Optical Communication System

SHERMAN KARP, MEMBER, IEEE, AND ROBERT M. GAGLIARDI, MEMBER, IEEE

REFERENCE: Karp, S., and Gagliardi, R. M.: THE DESIGN OF A PULSE-POSITION MODULATED OPTICAL COMMUNICATION SYSTEM, NASA/Electronics Research Center, Cambridge, Mass. 02139, and University of Southern California, Los Angeles, Calif. 90007. Rec'd 7/30/69; revised 5/9/69. Paper 69TP47-COM, approved by the IEEE Radio Communication Committee for publication without oral presentation. IEEE TRANS. ON COMMUNICATION TECHNOLOGY, 17–6, December 1969, pp. 670–676.

ABSTRACT: In recent literature the advantages of an idealized narrow-width pulse-position modulated (PPM) optical communication system, using coherent sources and direct photodetection, have been shown. In this paper the practical design of such an operating PPM link is considered. System performance in terms of error probabilities and information rates is derived in terms of key parameters, such as power levels, number of PPM signals, pulse width, and bandwidths. Both background radiation and receiver thermal noise are included. Design procedures utilizing this data are outlined. Whenever possible, optimal design values and parameter tradeoffs, in terms of maximizing information rate or minimizing transmitter power, are shown. The effect on performance of photomultipliers and their inherent statistics is also presented. Although the basic analysis is derived in terms of photon "counts," the necessary system optics equations are introduced to allow for overall optical hardware design. The primary underlying assumption is that synchronization is maintained at all times between the transmitter and receiver.

Introduction

WITH THE development of coherent sources in the optical region of the spectrum, there has been an increasing interest in the design of optical communication systems [1]–[6]. The direct detection of optical radiation is presently restricted to photodetection surfaces, for which it has been shown that the released electrons obey Poisson statistics [7], [8]. In this case investigators have shown the advantages of using narrow-width pulse-position modulation (PPM) as the principal mode of communication [3], i.e., coding information into one of M possible signals and transmitting it as a pulse of optical energy placed in one of M adjacent time intervals. It had been shown [1]–[3] that idealized versions of such systems optimize performance in terms of both various "distance" criteria and overall error probability, for cases of most interest. Therefore, in this paper we shall present procedures for the practical design of an M-ary low "duty cycle" PPM optical communication systems. In particular, performance characteristics in terms of key system parameters will be derived, with emphasis on hardware limitations and interference effects.

Consider the PPM optical communication system shown in Fig. 1. The transmitter is a monochromatic optical source operating at a fixed frequency. Information is sent

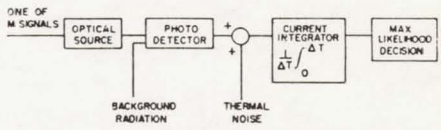


Fig. 1. Maximum-likelihood processor.

by transmitting one of M signals as a pulse of optical energy at the same frequency, located in one of M adjacent time intervals, each of which is ΔT seconds wide. We assume that complete synchronization is maintained between the transmitter and receiver at all times, i.e., time coherent operation. The optical receiver detects the transmitted signal by attempting to determine the optical energy in each possible time slot, then selecting the signal which corresponds to the maximal energy. In direct photodetection this is equivalent to "counting" the number of released electrons in each ΔT interval. Background radiation entering the photodetector acts as erroneous energy, causing signaling errors. In practical systems photomultipliers are often used to afford an improvement in photodetection (i.e., a gain in numbers of released electrons) but unfortunately often behave randomly, complicating system design. In addition, additive thermal noise may occur after photodetection, which tends to cause further errors in signal decisioning. Both of these latter effects shall be considered subsequently.

As optical radiation impinges upon a photodetecting surface, a series of electrons are released; each produces a current pulse $eh(t - t_m)$, where e is the electron charge, t_m is the time of release, and $h(t)$ represents the current motion. The function $h(t)$ is pulse like, having a time width roughly equal to the inverse of the photodetector bandwidth. We assume that $h(t)$ is identical for all electrons, and that the area of $h(t)$ is normalized to unity. In the absence of thermal noise the output voltage across a resistor R of the normalized current integrator, when sampled after ΔT seconds of integration, is then

$$v = (Ke/\Delta T)R \tag{1}$$

where K is the number of electrons released during the ΔT-second integration period. This result neglects "end effects;" that is, it assumes that $h(t)$ can be considered an "impulse" function with respect to the integrating time ΔT. Thus the integrator sample is proportional to the number of electrons released in the preceding ΔT time interval and therefore "counts" electrons. The average number of electrons produced in the time interval ΔT as a result of the received radiation from the transmitter is denoted K_S and is proportional to the received trans-

mitter energy, i.e.,

$$K_S = \eta P_R \Delta T / hf \tag{2}$$

where η is the photodetector efficiency, P_R the received peak optical signal power in watts, and f the mode (transmitter) frequency, and where $h = 6.6 \times 10^{-34}$ joules per second. The received power can be related to the system optics by

$$P_R = (d/\theta r)^2 P_T L \times 10^{-4} \tag{3}$$

where d is the optical diameter in centimeters, r the range in meters, θ the divergence of the antenna in radians, P_T the average transmitter power, and L the transmitting optics loss. The peak transmitter power can be converted to the average transmitter power by dividing it by the number of signaling intervals M.

Similiarly, we denote by K_N the average number of electrons in a time interval ΔT produced by the background radiation received by the optical collector. Thus

$$K_N = \eta P_N \Delta T / hf \tag{4}$$

where now P_N is the received background radiation average power. This is generally written as

$$P_N = N_\lambda A_R L_o \Omega \Delta\lambda \tag{5}$$

where N_λ is the background spectral radiance [watts per area (solid angle) bandwidth], L_o the optical loss, A_R the area of the collector, Ω the resolution of the receiver (solid angle), and $\Delta\lambda$ the optical bandwidth.

Note that with P_N held constant, the average number of noise electrons K_N is proportional to ΔT. This clearly indicates the advantage gained by low "duty cycle" operation, i.e., by the use of signal intervals which are as narrow as possible to decrease the amount of interfering radiation. The minimum value for ΔT, however, is approximately $1/\Delta\lambda$, for then the assumption of fixed noise power P_N is no longer valid. (For optical filters on the order of 5 Å, ΔT widths of 10^{-11}–10^{-12} seconds are feasible.)

The number of electrons counted in a signaling interval (i.e., an interval ΔT containing transmitter energy) is a Poisson random variable whose average value is $K_S + K_N$. For nonsignaling intervals the average value is K_N. We have tacitly assumed that the system is synchronized, i.e., that the integration occurs exactly during the ΔT data intervals. The maximum-likelihood processing corresponds to counting the number of electrons in each of the M intervals and selecting the interval with the largest count as the proper PPM signal. Allowing for likelihood draws (in which case we make a random selection among the drawees), the probability of making a correct decision is

$$P_D = \sum_{r=0}^{M-1} 1/(r+1) \left[\begin{array}{l}\text{i.e., the probability that the}\\ \text{correct interval count equals } r\\ \text{other interval counts and ex-}\\ \text{ceeds the remaining } M-r+1\end{array}\right].$$

Temporarily neglecting the additive thermal noise and taking into account all the ways in which the correct interval count can equal r other interval counts, we have

$$P_D = \sum_{x=1}^{\infty} \sum_{r=0}^{M-1} \frac{1}{r+1} \binom{M-1}{r} \frac{(K_S + K_N)^x}{x!} \cdot \exp\left[-(K_S + K_N)\right] \left[\sum_{i=0}^{x-1} \frac{K_N{}^i}{i!} \exp(-K_N)\right]^{M-1-r} \cdot \left[\frac{K_N{}^x}{x!} \exp(-K_N)\right]^r + (1/M) \exp\left[-(K_S + MK_N)\right]. \tag{6}$$

Using the identity

$$\sum_{r=0}^{M-1} \frac{(M-1)!}{(r+1)!(M-1-r)!} A^{M-1-r} B^r = \frac{A^{M-1}}{M(B/A)} \left[(1 + B/A)^M - 1\right]$$

we can rewrite (6) as

$$P_D = \sum_{x=1}^{\infty} \left\{ \frac{(K_S + K_N)^x}{x!} \exp\left[-(K_S + K_N)\right] \cdot \left[\sum_{i=0}^{x-1} \frac{K_N{}^i}{i!} \exp(-K_N)\right]^{M-1} \left[\frac{(1+a)^M - 1}{Ma}\right] \right\} + (1/M) \exp\left[-(K_S + MK_N)\right] \tag{7}$$

where

$$a = K_N{}^x \Big/ x! \sum_{i=0}^{x-1} \frac{K_N{}^i}{i!}.$$

This result is amenable to computation and can be used in system design to obtain performance characteristics for M-ary operation with fixed parameter values. An exemplary plot is shown in Fig. 2, in which error probability $P_E = 1 - P_D$ is plotted versus M for fixed values of K_S and K_N. The results show the degradation in system performance as M is increased, which can be attributed to the increase in the likelihood draws as the number of intervals increases. Note that P_E depends upon both K_S and K_N and not simply upon their ratio, so that a complete catalog of P_E curves is required to handle all design conditions [11].

Previously, it was stated that the signal intervals ΔT should be as narrow as the optical bandwidths allow. This fact can be shown quantitatively by examining P_E as a function of ΔT, assuming P_n fixed. This is shown, for example, in Fig. 3, where $M = 2$, and where P_n is constrained such that the average electron noise count in an interval T_0 is 10. The probability of error is plotted versus $\Delta T/T_0$, the duty cycle of the transmitter. Since K_S is held fixed throughout each curve, the transmitter peak power must necessarily increase proportionally, as is obvious from (2). Note that the error probability decreases monotonically as ΔT decreases. The minimum values at

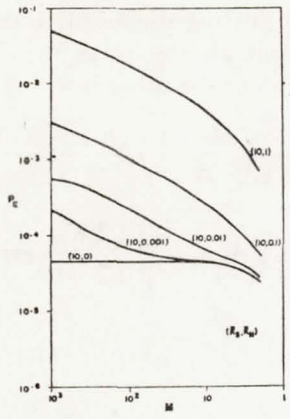


Fig. 2. Error probability versus M.

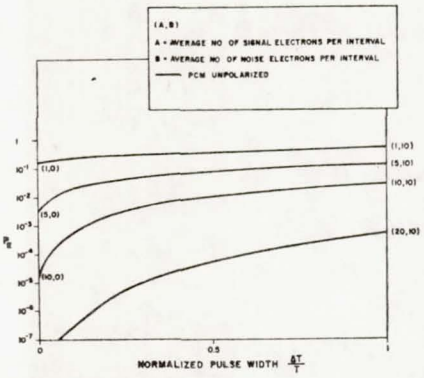


Fig. 3. Error probability versus normalized pulse width.

$\Delta T = 0$ are shown, but operation at these values implies infinite optical bandwidth. This minimum value is precisely the probability that a zero count occurs in the signaling interval (which is also the noise count of the nonsignaling intervals), and that the receiver randomly selects incorrectly. It is also interesting to note that the pulse width ΔT is not particularly significant at low values of K_S. As K_S increases, however, the duty cycle begins to play a paramount role in the resulting error probability. Thus attempts to increase bandwidth will have a direct payoff in system operation.

It is generally tempting for communication engineers to base system design in terms of signal-to-noise ratios (SNR). Typically, this is defined as the ratio of the squared average signal electron count to the variance of the count with background noise [5]. In terms of our previous notation this becomes.

$$S/N = K_S^2/(K_S + K_N). \tag{8}$$

To indicate the difficulty in basing design purely upon the SNR, consider the following comparison. First, let $K_S = 10$ and $K_N = 10$; then $S/N = 5$ and, from Fig. 2, $P_E = 3 \times 10^{-2}$. On the other hand, consider the case when $K_S = 5$ and $K_N = 0$. Again $S/N = 5$, but now $P_E = 3 \times 10^{-3}$. Thus with less signal energy we have improved the detection an order of magnitude with the same SNR. This is due to the fact that P_E in (7) depends upon signal energy K_S *and* noise energy K_N and not only

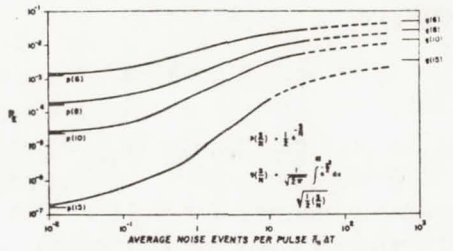


Fig. 4. Probability of error for fixed SNR versus noise background.

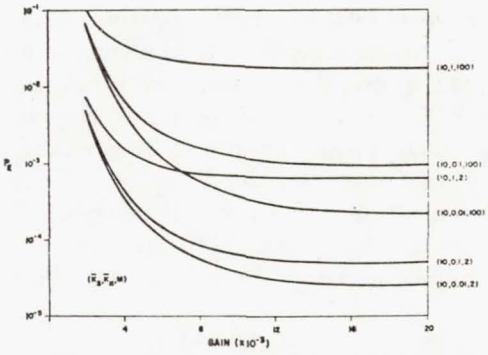


Fig. 5. Probability of error for ideal photomultiplier.

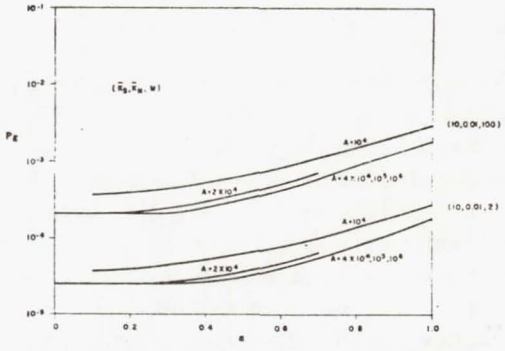


Fig. 6. Probability of error for photomultiplier model.

upon the ratio. It is precisely this point that distinguishes the Poisson detection problem from the analogous problem of transmitting one of M orthogonal signals over a Gaussian additive channel of equal noise power.

It is also of interest to determine how the SNR and error probability are related, in general. By solving for K_S in (8) one obtains

$$K_S = \tfrac{1}{2}(S/N)\{1 + [1 + 4K_N/(S/N)]^{1/2}\} \tag{9}$$

which can be inserted into (7). We can then plot P_E versus K_N for fixed values of S/N and M, as shown in Fig. 4. For large values of K_N, P_E approaches erf $(S/2N)$, which is identical to the error probability of orthogonal coherent signaling in Gaussian noise with a SNR of S/N. Since this approach is asymptotic from below, the Gaussian case will always be inferior to the Poisson case for the same SNR. This point is also in agreement with results showing the equivalence of optimum Poisson process-

ing and optimum Gaussian processing for large values of background noise power [1], [4]. Notice again that to maintain the same SNR, K_S must be increased. Lastly, we note that even if noise background is negligible, i.e., $K_N \to 0$, the signaling error probability does not go to zero but rather approaches

$$\lim_{K_N \to 0} P_E = \tfrac{1}{2} \exp(-K_S). \tag{10}$$

This is the same as the minimum values (at $\Delta T = 0$) shown in Fig. 3.

Effect of Thermal Noise and Photomultipliers

So far in our analysis we have neglected the effect of additive thermal noise, which adds a random variable to the integrator output sample of (1). This complicates our original assertion that the receiver counts exactly the number of electrons at the photodetector output. Since this is the crux of the maximum-likelihood direct detection system, it would be worthwhile to investigate this problem in more detail. Suppose, for example, that we consider the current resulting from the flow of a signal photoelectron during a time interval $\Delta T = 10^{-9}$ seconds. The sample value at the integrator output across a 50-Ω resistor from this electron would be 8×10^{-9} volts. Now if the receiver operated at room temperature, the thermal additive noise would contribute an integrator noise voltage whose root mean square (rms) value is approximately 28×10^{-6} volts ($R = 50\Omega$, temperature = 300°). Clearly, the count of a single electron could not easily be made in such a poor signal and noise condition. Photomultipliers exist, however, which effectively amplify the current effect of each photoelectron, resulting in a larger photoelectron count at the integrator output. Let this amplification factor be A, so that each photoelectron contributes a current value $Ae/\Delta T$ at the time of sample. We would like to determine the effect of A on P_D in (7) when Gaussian white thermal noise of one-sided spectral level N_0 is added at the integrator input. If we assume that each electron receives the same photomultiplier gain A, then the sample value due to the photodetector output

$$v = (KeA/\Delta T)R \tag{11}$$

where K is again the number of electrons produced during the interval ΔT. With Poisson statistics for the electron count during the ΔT interval, the probability density of v is then

$$P(v) = \sum_{j=0}^{\infty} [(\bar{K}^j/j!) \exp(-\bar{K})]\delta(v - jeA\beta R) \tag{12}$$

where $\delta(x)$ is the Dirac delta function, $\beta = 1/\Delta T$, and $\bar{K}$ is the average value of K. The thermal noise is integrated by the integrator and adds to the integrator sample v a random variable that is Gaussian distributed with zero mean and variance $N_0\beta$. Thus the total integrator sample value z after ΔT seconds of counting has a probability density obtained by convolving the Gaussian density with the discrete density in (12), yielding

$$P_z(z,\bar{K}) = \sum_{j=0}^{\infty} [(\bar{K}^j/j!) \exp(-\bar{K})]G(z,jAe\beta R,N_0\beta) \tag{13}$$

where $G(a,b,c)$ denotes a Gaussian density in the variable a with mean b and variance c. Observe that the sample probability densities are now continuous densities, and that the probability of equal sample values occurring is zero; that is, there is a zero probability of likelihood draws. Now the average count $\bar{K}$ is $K_S + K_N$ when a signal is present in the ΔT interval and is K_N when the signal is absent. Therefore, the probability of a correct decision is simply the probability that the observable z after the correct interval exceeds the observable z after the $M - 1$ remaining intervals. Hence

$$P_D = \int_{-\infty}^{\infty} dz\, p_z(z,K_S + K_N) \left[\int_{-\infty}^{z} p_z(y,K_N)\, dy\right]^{M-1}. \tag{14}$$

This can be written more compactly, as

$$P_D = \sum_{j=0}^{\infty} \frac{(K_S + K_N)^j}{j!} \exp[-(K_S + K_N)] \cdot \int_{-\infty}^{\infty} dz\, G(z,\mu,\sigma^2)\psi^{M-1}(z) \tag{15}$$

where

$$\psi(z) = \frac{1}{2}\left\{1 + \sum_{j=0}^{\infty}\left[\frac{K_N{}^j}{j!} \exp(-K_N)\right] \operatorname{erf}\left(\frac{z - j\mu}{(2\sigma)^{1/2}}\right)\right\}$$

$$\mu = A\epsilon\beta R$$

$$\sigma^2 = N_0\beta$$

and erf (x) is the error integral. Equation (15) has been evaluated for several values of K_S and K_N and is shown in Fig. 5 with $\beta = 10^9$ Hz and N_0 corresponding to a noise temperature of 300° with a 50-Ω load. The asymptotic values for large photomultiplier gains are precisely the values obtained by (7). At low gains, however, the thermal noise becomes the dominating source of error, and the probability of error increases rapidly. Note that to overcome the thermal environment a photomultiplier gain of about 10^4 is necessary for all the operating conditions shown. For other thermal environments one can use, as a rule of thumb, $A > 600(\text{temperature})^{1/2}$, obtained by directly scaling the foregoing results.

In the previous analysis we have assumed that the multiplier gain was a constant; that is, it was the same for all photoelectrons. In practice, however, the gain itself is generally a random variable [10] with a variance or "spread" which is usually taken as a percentage of the mean gain. We would now like to recompute P_D under this situation. If we let A_i be the electron gain of the ith released photoelectron, then the integrator sample

value for K electrons is

$$v = \sum_{i=1}^{K} A_i e\beta R \tag{16}$$

where each A_i is an independent random variable with probability density $P_A(X)$. The probability density of v is then obtained from

$$P(v) = \sum_{K=0}^{\infty} p(v/K)p(K) \tag{17}$$

where $p(v/K)$ is the conditional density of v, given K photoelectrons. Hence from (16)

$$p(v/K) = (1/e\beta R)^K [P_A(v/e\beta R) \underset{K-1}{\otimes} P_A(v/e\beta R)] \tag{18}$$

where $\otimes_K$ denotes K-fold convolution. Let us assume that $P_A(X)$ is Gaussian with mean A and standard deviation $\alpha A/2$, where $0 \leq \alpha \leq 1$ represents a percentage of the gain. Then (18) is a Gaussian random variable with mean $KAe\beta R$ and variance $K[(\alpha A/2)e\beta R]^2$ so that (17) becomes

$$P(v) = \sum_{i=0}^{\infty} (\bar{K}^i/i!) \exp(-\bar{K}) G\{v, iAe\beta R, i[(\alpha A/2)e\beta R]^2\}. \tag{19}$$

If one adds the sample contribution due to the thermal noise, then the sample z has a probability density

$$P_s(z,\bar{K}) = \sum_{i=0}^{\infty} (\bar{K}^i/i!) \cdot \exp(-\bar{K}) G(v, iAe\beta R, i[(\alpha A/2)e\beta R]^2 + N_0\beta). \tag{20}$$

Note that (20) is identical to (13), except for the variance terms in the Gaussian densities, and that it specializes to (13) for $\alpha = 0$. Hence the probability of detection is given exactly by (14), with the variance σ^2 replaced by this variance. The resulting error probabilities are shown in Fig. 6 as a function of the spreading parameter α, using parameters as in Fig. 5. Although the results vary somewhat as a function of the signal and noise, one observes that for mean gains between 10^4 and 10^5 the error probabilities obtained earlier (Fig. 2) are valid with gain spreads as high as 30 to 40 percent. Even with spreads as high as 70 percent, the results indicate only about a factor of 2 increase in error probability. The primary conclusion, then, is that with suitably high-mean photomultiplier gains, system error probabilities can be basically divorced of additive thermal noise effects. In this case the error curves plotted in Fig. 2 represent the overall system error probabilities.

The previous results also imply that a device average gain-to-spread ratio should be as large as possible for best operation. Consider an idealized photomultiplier characterized as a Poisson branching process [10]. Every photoelectron emitted from the photoemissive surface impinges on the first stage of the device and releases K secondary electrons which are Poisson distributed in number with parameter δ_1. The secondary electrons are then focused on the second stage where the same effect occurs. This process continues on through n stages, resulting in a large electron flow at the anode for each photoelectron emitted. The distribution of electrons at the anode output is quite complicated but is unimodal and quite easily fitted with a Gaussian distribution, as we have already done. In our calculations of error probability the two important parameters emerging were the mean gain of the device A and the mean gain to rms gain ratio Γ. For the idealized photomultiplier the two parameters can be calculated in a straightforward manner and are, in our previous notation,

$$A = \prod_{i=1}^{n} \delta_i$$

$$\Gamma = \frac{2}{\alpha} = \left(\frac{\delta_1}{(1 + 1/\delta_2 + 1/\delta_2\delta_3 + 1/\delta_2\delta_3\delta_4 + \cdots)} \right)^{1/2}.$$

Notice that Γ is almost completely characterized by the first-stage gain δ_1 (a minor contribution is also made by the second stage). Thus we can relate α directly to the gain of the first stage by

$$\alpha \cong 2/\delta_1^{1/2}.$$

For photon counting, Fig. 6 indicates that $\alpha < 0.4$, or that

$$2/\delta_1^{1/2} < 0.4 \Rightarrow \delta_1 > 25.$$

To take into account the effects of the remaining $n - 1$ stages, assume $\delta_2 = \delta_3 = \cdots = \delta_n = \delta$. Then

$$\sum_{i=0}^{n} (1/\delta)^i = \frac{1 - (1/\delta)^{n+1}}{1 - (1/\delta)} < \frac{1}{1 - (1/\delta)} = \frac{\delta}{\delta - 1}.$$

Therefore, δ_1 should be increased by $\delta/(\delta - 1)$. Typically $\delta = 4$, so that

$$\delta_1 > 25 \cdot \tfrac{4}{3} = 33.$$

Information Rate of a PPM System

We have so far analyzed only one aspect of system performance, i.e., error probabilities. The actual information rate that the link achieves is another important design consideration. As stated, the transmitter sends optical energy in one of M time intervals, which is ΔT seconds wide, thereby transmitting one of M possible signals in $M\Delta T$ seconds, or at a rate $\log_2 M/M\Delta T$ bit/s. The receiver correctly determines the true signal with probability $1 - P_E$ and is in error with probability P_E. Because of symmetry the erroneous signal may be equally likely interpreted as any of the $M - 1$ incorrect signals. Thus the overall channel may be depicted as an M-ary symmetric channel, in which each of the M possible transmitter signals is converted to itself with probability $1 - P_E$ and converted to each other signal with probability $P_E/(M - 1)$. The information rate for such a channel

is known to be

$$H = \frac{\log_2 M + P_E \log_2 [P_E/(M-1)] + (1-P_E)\log_2 (1-P_E)}{M\Delta T}. \quad (21)$$

For convenience we shall denote this as

$$H = C(K_S,K_N,M)/M\Delta T \quad (22)$$

to emphasize the dependence of the numerator on the stated parameters. By using (22) and the families of error probability curves as in Fig. 2, the rate H can be evaluated by straightforward substitution. Although specific curves for such a computation are not shown here, it suffices to note that if K_S and K_N are such that $P_E < 10^{-1}$, then (22) is, to a good approximation,

$$H \approx (1 - P_E)[(\log_2 M)/M\Delta T]$$
$$= (\log_2 M)/M\Delta T - P_E[(\log_2 M)/M\Delta T]. \quad (23)$$

If we interpret the rate H as the source rate minus the equivocation of the channel, then the PPM optical system behaves approximately as if a source rate of $\log M/M\Delta T$ is passed into a channel of equivocation $P_E \log M/M\Delta T$. As noted in (10), even if $K_N \to 0$ (no background interference), $P_E \to \exp(-K_S)/2$, so that the equivocation is not due entirely to the background noise.

The use of (22) and the previous equations are helpful in determining the rate, given operating parameters. However, the converse design problem, which is to determine particular parameter values that achieve a desired rate, is not so straightforward. This is due to the fact that the rate is a somewhat complicated function of the parameters. We shall consider here two aspects of this design problem that have practical application under certain operating conditions. First, the word period $T = M\Delta T$ is held fixed while the information bandwidth $\beta = 1/\Delta T$ is allowed to vary; and second, the system bandwidth β is held fixed while the word period is allowed to vary. In both cases we are interested in the relationship between the rate H and the transmitter parameters K_S and M, assuming that the noise power is held fixed.

Fixed Work Period

We assume here that ΔT is allowed to vary with M so as to maintain $T = M\Delta T$ constant. Thus the system "squeezes" more signals into the T-second period as M increases. The resulting rate is then

$$H = C(K_S,K_{NT}/M,M)/T \quad (24)$$

where K_{NT} is the noise energy in T. Thus the rate depends only upon the numerator of (22). With K_S fixed, increasing M increases the source rate, but the error probability also increases and eventually reaches an asymptotic value of

$$P_E = \left(1 + \frac{K_S[K_{NT} - 1 + \exp(-K_{NT})]}{K_{NT}}\right)\exp(-K_S)$$

for large M. The resulting system rate increases, to within a constant of the entropy of the alphabet, $\log_2 M/T$. Therefore, it is clear that if the bandwidth is expendable, one will always increase the system rate for large M by increasing M. In a practical system this implies that one should operate with as wide a bandwidth as possible to fully exploit the capability of the PPM system. We are therefore led naturally to consider the design of a system for an arbitrary rate H, when the full bandwidth $(1/\Delta T)$ of the system is limited.

Fixed Bandwidth

In this case ΔT is held constant (thereby fixing the noise energy K_N in ΔT) so that both the numerator and denominator in (22) depend upon M, and the rate degrades quickly as M increases due to the $\log M/M$ dependence. A given rate, e.g., H_0, may be obtained by many different combinations of K_S and M. Analytically, these equivalent operating points may be obtained graphically by noting that they are the values for which the numerator $C(K_S,K_N,M)$, considered as a function of M, intersects the straight line $H_0\Delta TM$. By plotting these functions for various K_S their intersection will identify (K_S,M) pairs which achieve the rate H_0. One may then decide on a particular operating point by invoking suitable design criteria. For example, one may select the smallest M from among the candidate pairs, which then minimizes the word period $T = M\Delta T$. Alternatively, one may choose to minimize the average transmitter power per information bit, which is proportional to K_S/C. In the latter case, therefore, one would select the operating pair (K_S,M) for which K_S/C is minimal. An application of this procedure is given in the next section.

An Example: Real-Time Television from Deep Space

To illustrate the design procedures outlined in this paper, we will consider a television system which will transmit in real time from deep space. System parameter values will be chosen to allow us to use the previously derived data and will not always reflect the optimum values or current state of the art. We shall assume the following transmission parameters:

background	blue sky
distance	4×10^8 km
optical loss	50 percent
receiver diameter	16 meters (nondiffraction limited)
receiver temperature	300°
optical bandwidth	5 Å
quantum efficiency	20 percent
resolution	1 arc second
photomultiplier gain	$>10^5$
optical frequency	5000 Å
signal bandwidth	10^9 Hz.

For real-time television a rate of approximately 7×10^7 bit/s is required (corresponding to better than 2 samples

TABLE I

K_S	M	K_S/C
0.8	19	0.593
0.9	23	0.552
1.0	28	0.506
1.2	35	0.494
1.4	42	0.484
1.6	47	0.488
1.8	52	0.496
2.0	55	0.508
5.0	95	0.792

per information hertz and 7 bits per sample coding). Our objective is to determine design parameters for a PPM system that uses minimal average transmitter power and a bit error probability no greater than 10^{-4}. Using this list of parameters in (4) and (5) yields $K_N = 0.01$ as the noise background count. Following the discussion in the previous section we plot $C(K_S,0.01,M)$ as a function of M, and determine the intersection with the line $(H_0\Delta T)M = (7 \times 10^{-2})M$, yielding the tabulation in Table I. The minimal average transmitter power occurs when $K_S = 1.4$ and $M = 42$ which defines the PPM system. Using (2) and (3), the transmitter then requires 0.072 watts. The corresponding P_E can be obtained from (7) to get the word error probability (2.4×10^{-1}), for which the corresponding bit error probability is approximately $P_E/2 \simeq 10^{-1}$ [9]. Thus the bit error is larger than that desired and in fact will be further increased by the thermal noise, as evidenced by the data of Fig. 5. (This figure shows a slight increase in P_E for $K_S = 10$ and $M = 2$ at photomultiplier gains of 10^4, and one can conclude the situation will be worse for $K_S = 1.4$ and $M = 42$). Thus the minimal power condition is not sufficient to obtain the desired P_E for this example without coding.

To achieve the desired bit error probability, we note from Fig. 2 that a system with $K_S = 10$ and $M = 100$ yields bit error probabilities $P_E/2 \approx 10^{-4}$ at the same noise level. We would expect no appreciable degradation from thermal noise, and Fig. 6 indicates that only slight increases occur even with photomultipliers having spreading as high as 45 percent. The transmitter power, which no longer is minimal for the desired information rate, is found to be 0.25 watts, an increase of 5.4 dB over the minimal average conditions.

Conclusion

In this paper we have considered some design aspects of an optical M-ary PPM communication system using photon counters (photodetectors followed by current integrators) at the receiver. The system considered transmits monochromatic optical energy in one of M time intervals, and the receiver determines the photon count (i.e., the received energy) in each interval and performs a maximum-likelihood test to determine which signal is being received. Complete time synchronization is assumed to be maintained at all times. Performance characteristics are derived in terms of system parameters with both background radiation and thermal noise interference. In particular, it is shown that, unlike the case of pure Gaussian additive noise, system performance does not depend on a few key parameters but must be recomputed for different operating points. The important equations for deriving these characteristics are introduced. From these equations, design procedures are outlined which lead to best choices of transmitter power, numbers of signals, interval lengths, etc., in order to obtain desired error probabilities and information rates. In particular, we have indicated the requirements imposed upon the individual components and have delineated the strong and weak points in the overall system. No attempt has been made to estimate the cost or weight in building such a system. This in fact would be premature since the technology required is in its infancy and is undergoing swift and quite radical change. In addition, we have not included any discussion of the methods to maintain link synchronization nor any consideration of atmospheric effects other than a loss factor.

Acknowledgment

The authors would like to thank M. G. Hurwitz and H. Asai for providing computational assistance.

References

[1] B. Reiffen and H. Sherman, "An optimum demodulator for Poisson processes: photon source detectors," *Proc. IEEE*, vol. 51, pp. 1316–1320, October 1963.
[2] K. Abend, "Optimum photon detection," *IEEE Trans. Information Theory* (Correspondence), vol. 12, pp. 64–65, January 1966.
[3] R. M. Gagliardi and S. Karp, "M-ary Poisson detection and optical communications," *IEEE Trans. Communication Technology*, vol. COM-17, pp. 208–216, April 1969.
[4] C. W. Helstrom, "Quantum limitations on the detection of coherent and incoherent signals," *IEEE Trans. Information Theory*, vol. IT-11, pp. 482–490, October 1965.
[5] W. K. Pratt, "Binary detection in an optical polarization modulation communication channel," *IEEE Trans. Communication Technology* (Concise Papers), vol. COM-14, pp. 664–665, October 1966.
[6] S. Karp, "A statistical model for radiation with applications to optical communications," Ph.D. dissertation, Dept. of Elec. Engrg., University of Southern California, Los Angeles, Calif., 1967.
[7] B. M. Oliver, "Thermal and quantum noise," *Proc. IEEE*, vol. 53, pp. 436–454, May 1965.
[8] L. Mandel, "Fluctuations of light beams," in *Progress in Optics*, E. Wolf, Ed. New York: Wiley, 1963.
[9] A. J. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966.
[10] H. J. Gale and J. A. B. Gibson, "Methods of calculating the pulse height distribution at the output of a scintillation counter," *J. Sci. Inst.*, vol. 43, no. 4, pp. 224–228, 1966.
[11] S. Karp, M. G. Hurwitz, and R. M. Gagliardi, "Error probabilities for maximum likelihood detection of M-ary Poisson processes in Poisson noise," NASA, Tech. Note TN-D-4721, October 1968.

Sherman Karp (M'62), for a photograph and biography please see page 216 of the April, 1969, issue of this Transactions.

Robert M. Gagliardi (S'57–M'61), for a photograph and biography please see page 216 of the April, 1969, issue of this Transactions.

Reprinted by permission from IEEE TRANSACTIONS ON COMMUNICATION TECHNOLOGY
Vol. COM-17, No. 6, December 1969
pp. 670-676

On the Representation of a Continuous Stochastic Intensity by Poisson Shot Noise

SHERMAN KARP, MEMBER, IEEE, AND ROBERT M. GAGLIARDI, MEMBER, IEEE

Abstract—In many applications a Poisson shot noise (PSN) process is said to statistically "represent" its intensity process. In this paper an investigation is made of the relationship between a PSN process and its intensity, when the latter is a sample function of a continuous stochastic process. The difference of the moments and the mean-square difference between the two processes are examined. The continuity assumption on the intensity permits the development of a sequence of moment relationships in which the effect of the PSN parameters can be seen. The results simplify and afford some degree of physical interpretation when the component functions of the PSN are "rectangular," or when the intensity process does not vary appreciably over their time width. An integral equation is derived that defines the component function that minimizes the mean-square difference between the two processes. It is shown that a "degenerate" form of component function induces complete statistical equality of the two processes. The problem has application to optical communication systems using photodetectors.

I. Introduction

IN THIS PAPER a study is made of the relationship between a Poisson shot noise process and its inherent intensity when the latter is a sample function of a continuous stochastic process. The problem is of interest in certain applications where a shot noise process is used to "represent" the intensity process. When the intensity is a deterministic time function, the relations between the shot noise and its intensity are well known [1]–[3]. However, when the intensity is itself a stochastic process, the manner in which the statistics of the intensity and shot noise are related is somewhat vague.

The problem has primary application to optical communications where shot noise processes are generally accepted as models for the output of wide-band photodetectors. In such models, the intensity of the received radiation impinging on the detector surface becomes the intensity parameter of the shot noise. When the received intensity corresponds to a desired modulating process (e.g., if the optical transmitter is intensity modulated with the process) a question then arises as to the context in which the detector has "demodulated" the input radiation.

Manuscript received April 1, 1969; revised August 21, 1969. This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-004. This grant is part of the research program at NASA's Electronics Research Center, Cambridge, Mass.

S. Karp was formerly with the Department of Electrical Engineering, University of Southern California. He is presently with NASA Research Center, Cambridge, Mass. 02139.

R. M. Gagliardi is with the Department of Electrical Engineering, University of Southern California, Los Angeles, Calif. 90007.

Other applications of similar shot noise modeling involve radiation scattering and perturbation effects [4], [5].

II. Moment Differences

Let $P(t_1, t_2)$, $t_2 > t_1$ indicate the number of random events occurring in the time interval (t_1, t_2). The number of events is Poisson distributed over (t_1, t_2) if the probability of k events in (t_1, t_2) is given by

$$\left[\Pr P(t_1, t_2) = k\right] = \left[\int_{t_1}^{t_2} n(t)\,dt\right]^k \cdot \exp\left[-\int_{t_1}^{t_2} n(t)\,dt\right]/k! \tag{1}$$

where $n(t)$ is an integrable nonnegative function and is called the intensity of the events. A Poisson shot noise (PSN) process is then defined as

$$I(t) = \sum_{m=1}^{P(-\infty, t)} h(t - t_m) \tag{2}$$

where $h(t)$ is the component function of the process and $\{t_m\}$ is a sequence of independent random variables, each having probability density

$$p_{t_m}(t) = \frac{n(t)}{\int_{-\infty}^{\infty} n(t)\,dt}. \tag{3}$$

It is well known [1], [2] that the semi-invariants of the first-order density of the process $I(t)$, at any t, are given by

$$\lambda_n(t) = \int_{-\infty}^{\infty} h^n(t - x)n(x)\,dx. \tag{4}$$

When the intensity $n(t)$ is a sample function of a stochastic process, the semi-invariants are themselves random processes, and the statistics of the process $I(t)$ have a complicated relation to those of the process $n(t)$. By making use of some practical assumptions concerning $n(t)$ and $h(t)$, we can, nevertheless, derive some properties of this relationship.

Let $n(t)$ be a sample function from a continuous nonnegative real bounded stationary random process N. (By continuous we mean every sample function is everywhere continuous, almost surely. By bounded we mean at every t, $n(t)$ is a bounded random variable.) Let the component function $h(t)$ be nonnegative real, time limited to τ seconds (i.e., $h(t) = 0$ for all t outside $(0, \tau)$, and $h^k(t)$ is integrable over $(0, \tau)$ for all k. Then the continuity of the process N

allows us to apply the mean-value theorem for integrals[1] to rewrite the conditional (on N) semi-invariants as

$$\lambda_n = n(\hat{t}) \int_{t-\tau}^{t} h^n(t - x)\, dx \qquad t - \tau \leq \hat{t} \leq t \tag{5}$$

where we have dropped the t variable from the λ_n description for convenience. The first-order moments of the PSN process $I(t)$ can now be directly related to the process N. We write the general kth moment as

$$E[I(t)]^k = E(E[I^k(t)|N]) \tag{6}$$

where E is the expectation operator and $E(\cdot|N)$ is expectation conditioned upon N. Using the relation between semi-invariants and moments [7], the conditioned expectations in (6) can be obtained from (5). Then substituting into (6), and using the stationarity of the process N, yields the sequence of moment relations:

$$\begin{aligned} E[I(t)] &= E(N)\epsilon_1 \\ E[I^2(t)] &= E(N^2)\epsilon_1^2 + E(N)\epsilon_2 \\ E[I^3(t)] &= E(N^3)\epsilon_1^3 + E[n(\hat{t}_1)n(\hat{t}_2)]\epsilon_1\epsilon_2 + E(N)\epsilon_3 \\ &\vdots \end{aligned} \tag{7}$$

where $E(N^k)$ is the kth moment of the process N, $\hat{t}_1, \hat{t}_2, \cdots$, are points in $(t - \tau, t)$, and

$$\epsilon_n = \int_0^{\tau} h^n(x)\, dx. \tag{8}$$

The lack of a general term relating moments and semi-invariants prevents us from writing an expression for the general kth moment. However, it is evident that the kth moment can be cast in the form

$$E[I^k(t)] = E(N^k)\epsilon_1^k + D(k). \tag{9}$$

In general, $D(k)$ represents a summation of terms in which each term is of the form

$$C(k, m, q)E[n^i(\hat{t}_m)n^j(\hat{t}_q)]\epsilon_m\epsilon_q \tag{10}$$

where $i + j \leq k - 1$. The $C(k, m, q)$ are positive constants generated from the expansion in (7). Equation (9) represents a general expression relating the moments of the PSN process to the statistics of N, in compliance with the aforementioned assumptions. It is clear that this relation depends not only upon the moments of N, but also upon its second-order statistics as well. Note that the component functions $h(t)$ enter the equation through both $\{\epsilon_i\}$ terms and the $\{\hat{t}_i\}$ parameters.

An important result concerning the process signal-to-noise ratio can be stated from (9). If we define the process SNR as the ratio of the square of its mean to its variance, we have

$$\begin{aligned} (\mathrm{SNR})_I &\triangleq \frac{E^2[I(t)]}{\mathrm{var}\,[I(t)]} = \frac{[E(N)\epsilon_1]^2}{E(N^2)\epsilon_1^2 + E(N)\epsilon_2 - [E(N)\epsilon_1]^2} \\ &= \frac{E^2(N)/\mathrm{var}\,(N)}{1 + [E(N)\epsilon_2/\epsilon_1^2\ \mathrm{var}\,(N)]} < (\mathrm{SNR})_N. \end{aligned} \tag{11}$$

where $(\mathrm{SNR})_N$ is the signal-to-noise ratio of the intensity process N. Hence, the SNR of the PSN process is always less than that of its intensity process. We make this point mainly because the above definition of SNR is commonly used in assessing signal quality in communication system analysis.

For convenience, we can normalize (9). Define the intensity of the PSN process to be normalized by the factor ϵ_1, and denote the resulting PSN process by $I_0(t)$. That is, we consider the normalized PSN process $I_0(t)$ whose intensity process has sample functions

$$n_0(t) = \frac{n(t)}{\epsilon_1}. \tag{12}$$

For this case, (9) becomes

$$E[I_0^k(t)] = E(N^k) + D_0(k) \tag{13}$$

where $D_0(1) = 0$, $D_0(2) = E(N)\epsilon_2/\epsilon_1$, $D_0(3) = \{E[n(\hat{t}_1)n(\hat{t}_2)]\epsilon_2 + E(N)\epsilon_3\}/\epsilon_1$, etc. That is, $D_0(k)$ is a normalized form of $D(k)$ and, in general, contains terms identical to those of $D(k)$ in (10), divided by the factor ϵ_1^{i+j}. When written as in (13), $D_0(k)$ represents the difference between the moments of process N and the moments of the normalized PSN $I_0(t)$, the latter having intensity process N_0 in (12). If the PSN is to "represent" the process N in the kth moment, then $D_0(k)$ should be "small" compared to that moment. Consider, for example, the relation of the mean-square values of $I_0(t)$ and $n(t)$. In this case, $D_0(2) = E(N)\epsilon_2/\epsilon_1$ and minimization of $D_0(2)$ requires minimization of ϵ_2, which depends only upon $h(t)$. We may then inquire if there is one component function $h(t)$, $0 \leq t \leq \tau$ that will minimize ϵ_2 for a fixed ϵ_1. By straightforward application of calculus of variations, using Lagrange multipliers, we obtain the solution

$$\begin{aligned} h(t) &= \epsilon_1/\tau \qquad 0 \leq t \leq \tau \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{14}$$

That is, a "rectangular" function spread over the interval $(0, \tau)$. The solution in fact minimizes ϵ_n for all $n \geq 2$, but we must not hastily conclude that the rectangular component function minimizes moment differences $D_0(k)$ for all k. The rectangular function is, however, of interest not only for the above reason, but also because it simplifies (13), allowing further insight. Define $d = \epsilon_1/\tau$. Then $\epsilon_n = d^n\tau$, $\lambda_n = n(\hat{\tau})d^n\tau$, $\hat{\tau}_1 = \hat{\tau}_2 = \hat{\tau}_3 = \cdots = \hat{\tau}$, where $t - \tau \leq \hat{\tau} \leq t$. This allows us to substitute into (10) the identity $E[n^i(\hat{\tau})n^j(\hat{\tau})] = E(N^{i+j})$, for all i, j, which in turn

[1] The theorem referred to here is sometimes called the second mean-value theorem for integrals [6]. It states

$$\int_a^b f(x)g(x)\,dx = f(\theta)\int_a^b g(x)\,dx \qquad a \leq \theta \leq b,$$

and requires only the continuity of f, the positiveness of g, and the integrability of g and fg.

allows us to rewrite (13) as

$$E[I_0^k(t)] = E[N^k] + \sum_{i=1}^{k-1} C(k, i)E(N^{k-1})d^i$$

$$= E[N^k]\left[1 + \frac{\sum_{i}^{k-1} C(k, i)E(N^{k-1})d^i}{E(N^k)}\right]$$

$$\triangleq E[N^k][1 + D_0'(k)] \tag{15}$$

where $D_0'(k)$ denotes the right-hand bracketed term and the $C(k, i)$ are again positive constants (actually combinations of the previous $C(k, q, m)$ terms). Equation (15) allows us to conclude that the kth moment of the normalized PSN process, with rectangular component functions, will be approximately equal to the kth moment of the process N only *if* $D_0'(k) \ll 1$. The term $D_0'(k)$ can be evaluated knowing just the first k moments of N, except the constants $C(k, i)$ are not known in a general closed form.[3] Since the intensity process N is nonnegative, we can make use of well-known properties of absolute moments [8] to establish

$$E[N^k] \geq [E(N^{k-i})]^{k/k-i} = E[N^{k-i}][E(N^{k-i})]^{i/k-i}$$

$$\geq E(N^{k-i})[E(N)]^i. \tag{16}$$

Substituting into (15) then implies

$$D_0'(k) \leq \sum_{i=1}^{k-1} \frac{C(k, i)d^i}{[E(N)]^i} \tag{17}$$

and an upper bound on $D_0'(k)$ is seen to be inversely related to the average value of N. For

$$\frac{E(N)}{d} \gg [C(k, i)]^{1/i} \qquad \text{all } i \leq k,$$

(17) is approximately

$$D_0'(k) \leq \frac{k(k-1)d}{2E(N)} \tag{18}$$

where the bound is taken as simply the first term of the sum. This result implies that the kth moment of $I_0(t)$ is approximately equal to the kth moment of N, for all k for which

$$\frac{E(N)}{d} \gg \frac{k(k-1)}{2}. \tag{19}$$

Note that $E(N)/d = \tau E(N)/\epsilon_1 = \tau E(N_0) =$ (average number of occurrences in τ seconds). Thus, (19) essentially states that the "denseness" of the shot events (i.e., the average number of component functions occurring in the time interval of one component function) must be sufficiently large for moment representation. The right side of (19) serves as a rough rule of thumb for determining how large this denseness must be for approximate equality of the kth moment. It may be recalled [3] that for PSN with deterministic intensities, a condition of large number of occurrences is required before the PSN process loses its "discrete" nature. Equation (19) can therefore be interpreted as the statistical equivalent of this statement, i.e., the condition under which the PSN begins to take on the statistics of its intensity.

Note that (15), though strictly valid only for $h(t)$ rectangular, is also approximately valid if τ is much less than the time variations in $n(t)$, no matter what the component function is, since we would then be able to approximate with $n(\hat{t}_i) \approx n(t)$ for all $t - \tau \leq \hat{t}_i \leq t$. Thus, (16)–(19) are basically applicable for all applications in which the time variations in $n(t)$ do not change appreciably over the time width of the component functions. In fact, under the latter assumption, we may even remove the condition of stationarity on N, and (15) can be used with $E(N^k)$ replaced by $E[n^k(t)]$. From this, we can conclude that the above equations are valid for nonstationary continuous intensities and arbitrary component functions, so long as the component functions are sufficiently narrow in time width.

III. Degenerate PSN Processes

The previous results for rectangular component functions allow an additional interpretation. Let us examine a "degenerate" situation in which we let d, the rectangular function height, become arbitrarily small. We then note from (17) that as $d \to 0$, $D_0'(k) \to 0$ and

$$E[I_0^k(t)] \to E(N^k) \qquad \text{all } k.$$

Thus, all the moments of the degenerate process $I_0(t)$ become *identical* to the moments of the process N. Furthermore, since the process N is bounded, the moment principle of random variables [8] guarantees that, in fact, as $d \to 0$, the first-order probability density of $I_0(t)$ converges to that of the process N at every t. The limiting condition $d \to 0$ implies that the component rectangular functions degenerate to zero in amplitude, while the intensity of the $I_0(t)$ process, given by $n(t)/d\tau$, becomes infinite. That is, the component functions get "smaller," but their rate of occurrence, in forming the sum in (2), increases without bound. Loosely speaking, the functions become more "densely packed," and the PSN behaves more like a continuous process rather than a discrete process.

The actual manner in which the processes behave can be seen by investigating the conditional probability densities of $I_0(t)$ for $d \approx 0$. Since the semi-invariants λ_n are of order $0(d^n)$ for $n \geq 1$, we can establish that the conditional first-order probability density of $I_0(t)$ approaches a Gaussian density with mean $n(\hat{t}_1)$ and variance $dn(\hat{t}_2)$. For the case where the component functions are extremely narrow in time compared to the time variations of the process $n(t)$, we can approximate $n(\hat{t}_1) \approx n(\hat{t}_2) \approx n(t)$. Hence, given the sample function $n(t)$, the degenerate process $I_0(t)$ appears as a stochastic process with time-varying mean $n(t)$, and an additive Gaussian noise of variance $dn(t)$. The signal $n(t)$ and the noise are, of course, not independent, which distinguishes the true degenerate

[3] It can be seen, however, that $C(k, 1) = k(k-1)/2$, by noting the relation between semi-invariants and moments. We use this fact in (18).

shot noise representation from the more common signal-in-additive-noise representation.

We can, in fact, expand this notion, and show that the condition $d \to 0$ allows even stronger conclusion concerning the degenerate $I_0(t)$. The conditional correlation of the two random variables $I_0(t_i)$ and $I_0(t_j)$, $t_i \neq t_j$, is given by

$$E[I_0(t_i)I_0(t_j)|N] = E[\sum_m h(t_i - t_m) \sum_k h(t_j - t_k)|N]. \quad (20)$$

Expanding out the double summation, performing the expectation [see (35)], and applying the degenerate condition $d \to 0$, yields

$$E[I_0(t_i)I_0(t_j)|N] \underset{d\to 0}{\to} E[I_0(t_i)|N]\, E[I_0(t_j)|N]. \quad (21)$$

That is, the conditional correlation approaches the product of the conditional means as $d \to 0$. Thus, $I_0(t_i)$ and $I_0(t_j)$ are uncorrelated, and with the Gaussian condition proven earlier, are in fact independent. This argument can then be extended to prove the conditional mutual independence of a sequence of variables $\{I_0(t_i)\}$. The eventual conclusion is that

$$E\left[\prod_{i=1}^{r} I_0^{r_i}(t_i)\right] \underset{d\to 0}{\to} E\left[\prod_{i}^{r} n^{r_i}(t_i)\right] \quad (22)$$

where $\{r_i\}$ is an rth-order positive integer set. Thus, the rth-order moment of the degenerate process $I_0(t)$ approaches the corresponding rth-order moment of the process N, and the boundedness of N is again sufficient to guarantee convergence of the general rth-order probability density of $I_0(t)$ to that of N.

IV. Least-Mean-Square PSN Processes

The more meaningful results in the previous section were obtained for the case of rectangular or extremely narrow component functions, with degenerate amplitudes. In this section, we investigate the validity of these assumptions by approaching the problem from a different point of view. We attempt to determine if there exists an optimal component function $h(t)$, $0 \leq t \leq \tau$, that minimizes the mean-square difference between the shot noise process of (2) and an arbitrary nonnegative random process, V. We shall ultimately be interested in the particular case where the process V corresponds to the intensity of the shot noise, but initially we can allow some generality.

We formulate the problem as follows. Given the PSN process of (2) with intensity $n(t)$ as a sample function from a random process N. Let $v(t)$ be a sample function from random process V. Let N and V both be stationary nonnegative real processes with continuous cross correlation $R_{VN}(t)$, and autocorrelations $R_{NN}(t)$ and $R_{VV}(t)$, respectively. We define the mean-square difference

$$J = E[v(t) - I(t)]^2 \quad (23)$$

where $I(t)$ is again the PSN process of (2). We seek the nonnegative component function $h_0(t)$, $0 < t < \tau$, that minimizes J. Proceeding formally, we expand out (23) and compute the resulting expectations (the details are shown in the Appendix), yielding

$$J = E(V^2) - 2\int_{t-\tau}^{t} h(t - t_m)R_{VN}(t - t_m)\, dt_m + E(N)\int_{t-\tau}^{t} h^2(t - t_m)\, dt_m + \int_{t-\tau}^{t}\int_{t-\tau}^{t} h(t - t_m)h(t - t_k)R_{NN}(t_m - t_k)\, dt_m\, dt_k. \quad (24)$$

Now by straightforward application of the calculus of variations, subject to the constraint $h(t) = 0$ outside $(0, \tau)$, we derive the integral equation for the optimal component function $h_0(t)$, $0 < t < \tau$, that minimizes J. This is

$$R_{NV}(u) = E(N)h_0(u) + \int_{-\infty}^{\infty} h_0(u - s)R_{NN}(s)\, ds, \qquad 0 \leq u \leq \tau \quad (25)$$

$$= \int_{-\infty}^{\infty} h_0(u - s)[R_{NN}(s) + E(N)\,\delta(s)]\, ds, \qquad 0 \leq u \leq \tau \quad (26)$$

where $\delta(s)$ is the delta function. The equation has the form of a Wiener–Hopf equation. (Indeed, the problem could have been formulated in the context of mean-square filtering, since $h_0(t)$ can be regarded as a filter-impulse function operating upon a PSN process whose component functions are delta functions.) The problems in obtaining a general solution to (26) are developed in treatments on Wiener filter theory and need not be repeated here. The resulting mean-square difference, when the solution to (26) is substituted into (23), is

$$J_{\min} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{S_{VV}(\omega)[S_{NN}(\omega) + E(N)] - S_{VN}^2(\omega)}{S_{NN}(\omega) + E(N)}\right] ds \quad (27)$$

where $S_{VV}(\omega)$ is the Fourier transforms of $R_{VV}(t)$, etc. For our interest, we are concerned with the above development for the special case where $n(t) = v(t)/\epsilon_1$, as in (13). Under this assumption, $R_{NV}(t) = R_{VV}(t)/\epsilon_1$ and $R_{NN}(t) = R_{VV}(t)/\epsilon_1^2$. The corresponding equation for the optimal solution is

$$R_{VV}(u) = \int_{-\infty}^{\infty} h_0(u - s)\left[\frac{R_{VV}(s)}{\epsilon_1} + E(V)\,\delta(s)\right] d\omega = \frac{1}{\epsilon_1}\int_{-\infty}^{\infty} h_0(u - s)R_{VV}(s)\, ds + E(V)h_0(u), \qquad 0 \leq u \leq \tau. \quad (28)$$

The solution for $h_0(t)$ is not in general a rectangular function. This fact is not surprising, based upon our previous results, since we have removed the constraints in the derivation of (14). However, if we invoke the assumption that τ is much smaller than the time variations in $v(t)$ [which by the continuity of the correlation functions allows us to consider $R_{VV}(s) \approx R_{VV}(u)$ for all s in $(u-\tau, u)$], then (28) has only the trivial solution $h_0(u) = 0, 0 \leq u \leq \tau$. That this is indeed a minimizing solution can be verified

by substituting into (27) and noting that

$$J_{\min} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\epsilon_1 E(V) S_{vv}(\omega)}{S_{vv}(\omega) + \epsilon_1 E(V)}\, d\omega \to 0 \text{ as } \epsilon_1 \to 0. \tag{29}$$

The above solution for $h_0(u)$ corresponds to a form of degenerate PSN process introduced previously. Thus, the trivial solution to the desired mean-square-difference equation actually has physical meaning in the application presented here, as was discussed in Section III.

V. Conclusions and Applications

We have investigated some aspects of the relationship between a PSN process and its intensity, when the latter is a sample function of a continuous nonnegative real stochastic process. In particular, we examined the difference of the moments and the mean-square difference between the two processes. The continuity assumption on the intensity aided us in developing a sequence of moment relations that manifest the effect of the component function of the PSN. These results simplified, and afforded some degree of physical interpretation, when the component functions were taken as rectangles, or when the intensity did not vary appreciably over their time width. It has also been shown that a degenerate form of component function actually has meaning, and corresponds to the exact representation of the intensity by the PSN process. A principal conclusion of the paper is that a continuous nonnegative real process $v(t)$ can be "approximately" modeled by a PSN process by allowing $v(t)$ to be the intensity of the PSN, and properly assigning the component function. The results of the paper can be used for assessing the degree of approximation.

One particular application is in the field of optical communications, where photodetector outputs are modeled as PSN processes. In this case, the component function corresponds to a current "pulse" induced by the arriving photons. These current pulses have time widths inversely proportional to the detection bandwidth, on the order of 10^{-9} second. The intensity modulation has bandwidths on the order of 10^3–10^6 Hz. Thus, the assumption of little intensity variation over the time width of a component function is appropriate in this application. The average photon intensity of a coherent optical signal is given by P/hf, where P = the average transmitter power, h = Planck's constant, and f = frequency of the optical mode. The condition of the denseness of the shot noise (19) is therefore equivalent to the statement that P/hfB be sufficiently large, where B is the detector bandwidth. Since hf is often considered the intrinsic quantum noise spectral level, the parameter P/hfB appears as a power signal-to-noise ratio of the optical radiation. Hence, the detector output PSN process models the stochastic intensity modulation if the quantum signal-to-noise ratio is sufficiently high.

Appendix

The expansion of J in (23) is

$$J = E[v^2(t) - 2v(t)\sum_{m=1}^{p} h(t - t_m) + \sum_{m}^{p}\sum_{k}^{p} h(t - t_m)h(t - t_k)] \tag{30}$$

where we simplified $P(-\infty, t)$ to p. The expectation of the second term can be written

$$E\left[v(t)\sum_{m}^{p} h(t - t_m)\right] = E_{V,N}\left\{v(t)E_{t_m,p}\left[\sum_{m}^{p} h(t - t_m) \mid N, V\right]\right\} \tag{31}$$

where we have denoted by subscripts the variables over which we are averaging. By using (1) and (3), we have

$$E_{t_m,p}\left[\sum_{m}^{p} h(t - t_m)\right] = E(p)\int_{-\infty}^{\infty} h(t - t_m)p_{t_m}(t_m)\, dt_m = \int_{-\infty}^{\infty} h(t - t_m)n(t_m)\, dt_m. \tag{32}$$

Substituting into (31) then yields

$$E\left[v(t)\sum_{m}^{p} h(t - t_m)\right] = \int_{-\infty}^{\infty} h(t - t_m)E[v(t)n(t_m)]\, dt_m. \tag{33}$$

The third term in (30) contains terms for which $m = k$ and $m \neq k$. For the former

$$E\left[\sum_{m}^{p} h^2(t - t_m)\right] = E_N E_{t_m,p}\left[\sum_{m}^{p} h^2(t - t_m) \mid N\right] = E_N\int_{-\infty}^{\infty} h^2(t - t_m)n(t_m)\, dt_m = E(N)\int_{-\infty}^{\infty} h^2(t - t_m)\, dt_m. \tag{34}$$

For the $m \neq k$ terms, we have, by similar steps,

$$E\left[\sum_{m}^{p}\sum_{k}^{p} h(t - t_m)h(t - t_k)\right] = E_N\left\{E(p^2 - p)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t - t_m) \cdot h(t - t_k)p(t_m)p(t_k)\, dt_m\, dt_k\right\} = E_N\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t - t_m)h(t - t_k)n(t_m)n(t_k)\, dt_m\, dt_k \tag{35}$$

where we used the fact that $E(p^2 - p)/E^2(p) = 1$ for a Poisson random variable. The substitution of (33), (34) and (35) in (30) then yield (24).

References

[1] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–332, 1944.
[2] E. Parzen, *Stochastic Processes*. San Francisco: Holden-Day, 1962, p. 113.
[3] D. Middleton, *Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960, ch. 11.
[4] ——, "A statistical theory of reverberation and similar first-order scattered fields—Parts I and II," *IEEE Trans. Information Theory*, vol. IT-13, pp. 372–414, July 1967.
[5] S. Karp, R. Gagliardi, and I. Reed, "Radiation models using discrete radiation ensembles," *Proc. IEEE*, vol. 56, pp. 1704–1711, October 1968.
[6] R. Franklin, *A Treatise on Advanced Calculus*. New York: Wiley, 1958, ch. 4.
[7] H. Cramer, *Mathematical Methods of Statistics*. Princeton, N. J.: Princeton University Press, pp. 185–187.
[8] *Ibid.* pp. 174–179.

Reprinted by permission from IEEE TRANSACTIONS ON INFORMATION THEORY
Vol. IT-18, No. 1, January 1972, pp. 208-211

Photon Counting and Laguerre Detection

ROBERT M. GAGLIARDI

Abstract—**In this correspondence maximum-likelihood binary detection theory is applied to an incoherent optical system model employing photodetectors governed by Laguerre counting statistics. It is shown that a maximum-likelihood Laguerre detector corresponds to a count comparison over each signaling interval. Laguerre error probabilities are presented and compared with those for Poisson counting.**

Introduction

In optical communication systems binary data bits are often transmitted by sending light in one of two possible adjacent time intervals. When incoherent photodetection is used in each interval, the receiver is modeled as a counter of photons. The synthesis of the optimal receiver processing and its resulting performance therefore depend upon the statistics associated with this counting. In early work, the probability density of the counts was almost exclusively assumed to be Poisson, and the optimum receiver processing and performance [1]–[3], [8, pp. 207] were determined. Count statistics are in fact only conditionally

Manuscript received March 5, 1971; revised May 20, 1971. This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-104.

The author is with the Department of Electrical Engineering, University of Southern California, Los Angeles, Calif. 90607.

Poisson (e.g., see [4, pt. 1]), and the probability of k photoelectron counts occurring in a single spatial mode during a $(0,T)$ s counting interval is given by

$$P(k) = \int_0^\infty \frac{x^k}{k!} e^{-x} p(x)\, dx, \tag{1}$$

where $p(x)$ is the probability density of the random energy variable

$$x = \int_0^T |f(t)|^2\, dt$$

and $f(t)$ = detected optical field in a spatial mode at the photodetector. When the detected field is the sum of a deterministic signal field and Gaussian white, band-limited thermal noise of temperature μ and bandwidth B, and if the noise energy per each significant space-time mode is assumed equal, (1) is known to be [5]

$$P_L(k; E,N,D) = \frac{(N)^k}{(1+N)^{D+k+1}} \exp\left[-\frac{E}{1+N}\right] L_k^D\left[-\frac{E}{N(1+N)}\right]. \tag{2}$$

Here N is the average number of noise counts per time-space mode, given by Planck's formula $N = [\exp(hf/k\mu) - 1]^{-1}$, where h is Planck's constant, k is Boltzman's constant, and f is the optical carrier frequency. The parameter E is the average number of signal counts [signal energy over $(0,T)$ divided by hf], $L_k^D(\cdot)$ is the Laguerre polynomial

$$L_k^D(x) = \sum_{i=0}^{k} \binom{k+D}{k-i} \frac{(-x)^m}{m!}, \tag{3}$$

and $D = 2BT$ is the time-bandwidth product, often called the count dimension. Physically, $D+1$ is the number of temporal modes observed in a single spatial mode during the count interval. The probability in (2) is called a Laguerre probability, and we refer to the associated count statistics as Laguerre counting. Our objective now is the application of maximum-likelihood detection theory to an optical digital system governed by Laguerre counting. Earlier work in this area by Liu [9] and Helstrom [10] dealt with the case $D = 0$.

We may first digress to examine the conditions under which Laguerre counting can be replaced by Poisson counting. It has been shown [5] that if $N \to 0$, $D \to \infty$, in such a way that DN remains fixed, the Laguerre probability in (2) is asymptotic to the Poisson probability

$$P_p(k; E,DN) = \frac{(E+DN)^k}{k!} \exp[-E + DN)] \tag{4}$$

at every k. In a practical situation the condition $N \ll 1$ is generally true, since N is on the order of $10^{-7} - 10^{-6}$ for typical background noise sources and visible wavelengths. However, the dimension D depends upon the data rate being transmitted. A question then arises as to how large D should be in order to replace the Laguerre probability by the Poisson in analysis. A first-order condition can be determined by noting that if $N \ll 1$, the first two factors in (2) are, to a good approximation,

$$\left(\frac{1}{1+N}\right)^{D+1} \left(\frac{N}{1+N}\right)^k \exp[-E/(1+N)] \approx N^k \exp -(E+DN). \tag{5}$$

The Laguerre term in (3) can be rewritten by applying the asymptotic relation for the ratio of gamma functions [6, p. 15]

$$\frac{(D+k)!}{(D+j)!} = D^{k-j}\left[1 + \frac{(k-j)(k+j-1)}{2D}\right] + 0(|D|^{-2}). \tag{6}$$

This allows the Laguerre polynomial, for $N \ll 1$, to be written as

$$L_k^D\left[\frac{-E}{N(1+N)}\right] = \frac{1}{k!}\left(D + \frac{E}{N}\right)^k + \frac{1}{k!}\sum_{j=0}^{k}\left[\frac{(k-j)(k+j-1)}{2D}\right] \cdot \binom{k}{j} D^{k-j}\left(\frac{E}{N}\right)^j + 0(|D|^{-2}). \tag{7}$$

When (5) and (7) are substituted into (2), the Laguerre probability takes the form

$$P_L(k; E,N,D) = P_p(k; E,DN) + C(k,D,E,N) + 0(|D|^{-2}). \tag{8}$$

Here the function C appears as a first-order correction term to the Poisson probability and is due to the sum term in (7). If we apply the fact that

$$\frac{(k-j)(k+j-1)}{2D} \le \frac{k^2+k}{2D}, \qquad j \ge 0 \tag{9}$$

to (7), we can bound the correction term C in (8) by

$$C(k,D,E,N) \le \left[\frac{k^2+k}{2D}\right] P_p(k; E,DN). \tag{10}$$

Thus a bound can be placed on the first-order difference between Laguerre and Poisson probabilities. It implies that the probabilities may differ significantly over the tails, i.e., for large k. However, the form of (10) does allow a rough rule for bounding D. If, for example, the difference between Laguerre and Poisson probabilities is to be within a fraction γ of the Poisson value over a range of k, say $k \le k_0$, then D should at least satisfy

$$D \ge (k_0^2 + k_0)/2\gamma. \tag{11}$$

If D does not satisfy the above, then with certainty $(P_L - P_p)/P_p \ge \gamma$ for some $k \le k_0$. If D satisfies (11), the effect of higher order correction terms should be considered, although for $D \gg 1$, the first-order correction will predominate. (For example, a megabit system operating at 10 μ with a 1-Å optical filter will generate a D of about 400.)

Maximum-Likelihood Laguerre Detection

Consider an optical system in which two adjacent time intervals are used for binary signaling. Assume optical signals of equal energy are transmitted, additive background noise of constant energy level is encountered, and identical dimensions D exist during each counting interval. Let a binary one be represented by signal energy Ehf in the first interval and let a binary zero be represented by energy Ehf in the second interval. The receiver photodetects the received field (transmitted bit signal plus background noise) over each interval, producing photoelectron counts obeying the Laguerre statistics in (2). A decoder follows the photodetector and performs a maximum-likelihood test for deciding between the hypothesis of a binary one being received H_1 or a binary zero H_0. If k_i is the count over interval i, the decision is therefore based upon the observed count vector $\boldsymbol{k} = (k_1,k_2)$. If $p(\boldsymbol{k} \mid H_i)$ is the probability of $\boldsymbol{k}$ occurring when H_i is true, then the maximum-likelihood test corresponds to the decision rule

$$\text{decide} \begin{pmatrix} H_1 \\ H_0 \end{pmatrix} \quad \text{if } p(\boldsymbol{k} \mid H_1) \gtrless p(\boldsymbol{k} \mid H_0) \tag{12}$$

while an equilikely random decision is made when the $p(\boldsymbol{k} \mid H_i)$ are equal. For Laguerre counting, this corresponds to a comparison of the densities

$$p(\boldsymbol{k} \mid H_1) = P_L(k_1; E; N,D)P_L(k_2; 0,N,D) \tag{13}$$

and

$$p(\boldsymbol{k} \mid H_0) = P_L(k_1; 0,N,D)P_L(k_2; E,N,D). \tag{14}$$

Substituting from (2) and canceling common terms for a given $\boldsymbol{k}$ yield the equivalent comparison

$$L_{k_1}^D(-e)L_{k_2}^D(0) \gtrless L_{k_1}^D(0)L_{k_2}^D(-e) \tag{15}$$

or

$$\frac{L_{k_1}^D(-e)}{L_{k_2}^D(-e)} \gtrless \frac{L_{k_1}(0)}{L_{k_2}(0)}, \tag{16}$$

where $e = E/N(1+N)$. When $k_1 > k_2$ the function $L_{k_1}^D(-e)/L_{k_2}^D(-e)$ is monotonically increasing in e, which guarantees that the left side of (16) exceeds the right side. Similarly, when $k_2 < k_1$, the converse is true. This means the maximum-likelihood test in (12) is

equivalent to the test

$$\text{decide } \begin{pmatrix} H_1 \\ H_0 \end{pmatrix} \quad \text{if } k_1 \; (\gtrless) \; k_2,$$

$$\text{decide } H_1 \text{ with probability } \tfrac{1}{2} \quad \text{if } k_1 = k_2. \tag{17}$$

Thus, maximum-likelihood decoding with Laguerre counting requires only a count comparison over each interval.

When the maximum-likelihood test is implemented, the corresponding Laguerre error probability is given by

$$PE_L(E,N,D) = 1 - A \sum_{k_1=0}^{\infty} B^{k_1} L_{k_1}{}^{D}(-e) \sum_{k_2=0}^{k_1-1} B^{k_2} \binom{D+k_2}{k_2} - (A/2) \sum_{k_1=0}^{\infty} B^{2k_1} L_{k_1}{}^{D}(-e) \binom{D+k_1}{k_1}, \tag{18}$$

where $A = (1+N)^{-2(D+1)} \exp[-E/(1+N)]$ and $B = N/(1+N)$. The probability in (18) has been computed for several values of the parameters. Some typical plots are shown in Fig. 1 as a function of signal count E and crossplotted in Fig. 2 as a function of the counting dimension D. Fig. 2 is particularly useful for demonstrating the advantage of using narrow energy pulses (reducing T while keeping E fixed for a given optical bandwidth B) and thereby reducing error probability. This, of course, is simply a reiteration of the obvious fact that one should encounter as few noise modes as possible, while the signal energy should occupy the full available optical bandwidth. However, it can be seen that the advantage gained in PE is relatively small at low noise levels.

If Poisson counting had been assumed in each interval, with the same signal and total noise energies, the count probability in (4) would be used, and the error probability would instead be

$$PE_p(E,DN) = \sum_{k_1=0}^{\infty} \sum_{k_2=k_1}^{\infty} \frac{(DN)^{k_2}(E+DN)^{k_1}}{k_1!\,k_2!} \exp -[E+2DN] - \frac{1}{2} \sum_{k_1=0}^{\infty} \frac{[(E+DN)DN]^{k_1}}{k_1!\,k_2!} \exp -[E+2DN]. \tag{19}$$

The error probability in (19) is easier to compute and parameter studies have been extensively published [3], [7], [8, p. 215]. Some Poisson results have been superimposed in Figs. 1 and 2 to illustrate the difference between true error probabilities (Laguerre) and approximate error probabilities (Poisson). Two conclusions are immediately evident. The Poisson error probabilities are universally lower (more optimistic) than the corresponding Laguerre probabilities. Second, when $N \ll 1$, PE_p yields a fairly accurate approximation to PE_L, even if the dimension D is not particularly large. This fact appears to indicate that the discrepancies between $P_L(k)$ and $P_p(k)$ over the tails of the densities have little effect on error probability when the noise per mode is small. This is further emphasized in Fig. 3 in which E and DN are held fixed, while PE_L is plotted as function of D. The PE_p value for the same E and DN is shown as an asymptote. The PE_L curve approaches PE_p asymptotically from above as D increases, and the corresponding N decreases. In other words, for fixed signal and noise energies, PE_L is accurately predicted by PE_p if the noise energy is produced from a relatively low N value.

The magnitude of the difference between PE_L and PE_p depends upon the actual values of D, E, and N. This can be seen by investigating the behavior of the two functions at $D = 0$, which is the point at which the largest difference occurs. When $D = 0$, $PE_p(E,0) = \frac{1}{2} \exp[-E]$. The corresponding Laguerre limit can be determined by noting that for $D = 0$, (18) is

$$PE_L(E,N,0) = \left(\frac{1}{1+N}\right)^2 \left(\frac{1+2N}{2}\right) \exp[-E/(1+N)] \sum_{k=0}^{\infty} B^{2k} L_k{}^{D}(E'). \tag{20}$$

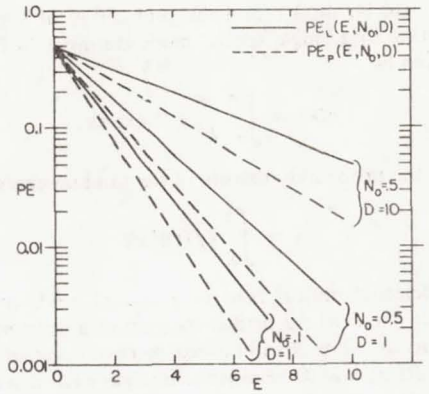


Fig. 1.

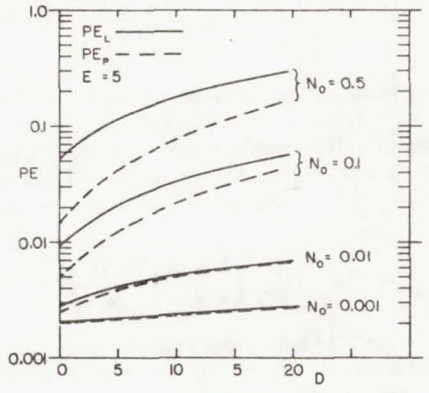


Fig. 2.

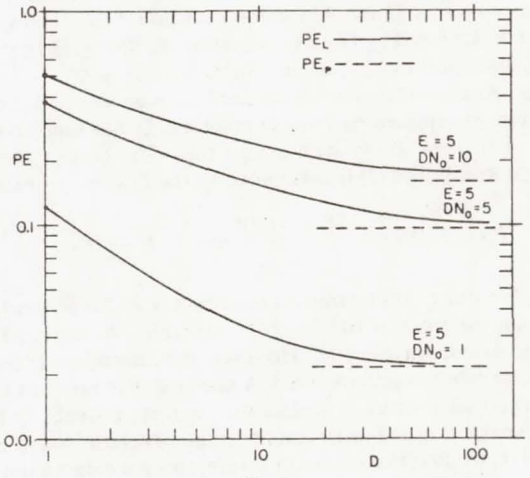


Fig. 3.

By applying a Laguerre identity [11, eq. (8.975)], and manipulating algebraically, the above becomes $PE_L(E,N,0) = \frac{1}{2} \exp[-E/(1+2N)]$. The ratio is then

$$\left.\frac{PE_L}{PE_p}\right|_{D=0} = \exp\left[\frac{2NE}{1+2N}\right]. \tag{21}$$

The above shows that the ratio of the two error rates depends explicitly upon the EN product. When expressed in this manner, one can append earlier statements and conclude that PE_p and PE_L are fairly close over the range of all $D \geq 0$ if both $N \ll 1$ and $NE \ll 1$.

In arriving at the above results it should be pointed out that the assumption of equal modal energy used in the derivation of (2) is strictly satisfied for white noise only for $D = 0$ and $D \gg 1$. For intermediate D values, say, $1 \leq D \leq 10$, this assumption is definitely

violated, and the PE_L points plotted using (2) for this range of D must be accepted as an assumed logical extension of the more accurate endpoints.

References

[1] B. Reiffen and H. Sherman, "An optimum demodulator for Poisson processes," *Proc. IEEE*, vol. 51, Oct. 1963, pp. 1316–1320.

[2] K. Abend, "Optimum photo detection," *IEEE Trans. Inform. Theory* (Corresp.), vol. 12, Jan. 1966, pp. 64–65.

[3] R. M. Gagliardi and S. Karp, "*M*-ary Poisson detection and optical communications," *IEEE Trans. Commun. Technol.*, vol. COM-17, Apr. 1969, pp. 208–216.

[4] S. Karp, E. L. O'Neill, and R. M. Gagliardi, "Communication theory for the free-space optical channel," *Proc. IEEE*, vol. 58, Oct. 1970, pp. 1611–1626.

[5] S. Karp and J. R. Clark, "Photon counting—A problem in classical noise theory," *IEEE Trans. Inform. Theory*, vol. IT-16, Nov. 1970, pp. 672–680.

[6] N. Lebedev, *Special Functions and Applications* (English transl.). Englewood, N.J.: Prentice-Hall, 1965.

[7] S. Karp *et al.*, "Error probabilities for Poisson detection," NASA Tech. Note TN-D-4721, Oct. 1968.

[8] W. K. Pratt, *Laser Communications*. New York: Wiley, 1969.

[9] J. W. S. Liu, "Reliability of quantum-mechanical communication systems," *IEEE Trans. Inform. Theory*, vol. IT-16, May 1970, pp. 319–330.

[10] C. W. Helstrom, "Performance of an ideal quantum receiver of a coherent signal of random phase," *IEEE Trans. Aerosp. Electron. Syst.* (Corresp.), vol. AES-5, May 1969, pp. 562–564.

[11] I. Gradsteyn and I. Ryzhik, *Tables of Integrals, Series and Products*. New York: Academic Press, 1965, p. 1037.

The Effect of Timing Errors in Optical Digital Systems

ROBERT M. GAGLIARDI, MEMBER, IEEE

Abstract—**The use of digital transmission with narrow light pulses appears attractive for data communications, but carries with it a stringent requirement on system bit timing. The effects of imperfect timing in direct-detection (noncoherent) optical binary systems are investigated using both pulse-position modulation (PPM) and on–off keying for bit transmission. Particular emphasis is placed on specification of timing accuracy and an examination of system degradation when this accuracy is not attained. Bit error probabilities are shown as a function of timing errors from which average error probabilities can be computed for specific synchronization methods. Of significance is the presence of a residual or irreducible error probability in both systems, due entirely to the timing system, which cannot be overcome by the data channel.**

I. Introduction

THE ABILITY to generate extremely narrow high-energy light pluses from a laser source has made the optical transmission of digital data extremely attractive for modern communications. This possibility has fostered an exhaustive exploration of optical communication systems, from both a theoretical and hardware point of view (e.g., see [1]). The use of digital transmission with narrow pulses, however, carries with it an extremely stringent requirement on system bit timing, i.e., time control of the system sampling and integration intervals during each data bit. For the most part, past analytical studies have assumed perfect system timing, and the degradation caused by timing errors in optical systems have been virtually ignored. In this paper, we investigate the effects of imperfect timing in a direct-detection (noncoherent) optical communication system, with particular emphasis on the specification of timing accuracy, and an examination of the system degradation when this accuracy is not attained.

Consider a general optical digital system as shown in Fig. 1(a). The system sends bits of information by transmitting bursts of optical energy. One of two possible methods are usually used for encoding the bits. In one, the system operates by transmitting a burst of energy in one of two T-second adjacent time intervals to encode a binary bit. This represents a two-level pulse-position modulated (PPM) mode of transmission and is known to be optimal under various criteria, when constrained in average transmitter power [2]. Thus, for example, the binary sequence 0110 would be transmitted by the optical waveform shown in Fig. 1(b), where the pulse represents a burst of optical laser energy. We have considered an energy pulse in the first interval to represent a binary one, and an energy pulse in the second interval to represent a binary zero. A second procedure is to use on–off keying, in which the transmitter uses an energy burst for a one, and transmits no energy for a zero. Thus, the waveform 0110 would be transmitted by the energy waveform in Fig. 1(c). Note that if T is the energy pulsewidth, then in PPM $2T$ is the bit interval and information is being transmitted at a rate $1/2T$ bit/s, while in on–off keying T is the bit interval and the rate is $1/T$ bit/s.

Paper approved by the Communication Theory Committee of the IEEE Communications Society for publication without oral presentation. This work was supported by NASA under Contract NGR-05-018-104. Manuscript received September 7, 1971; revised November 12, 1971.

The author is with the Department of Electrical Engineering, University of Southern California, Los Angeles, Calif. 90007.

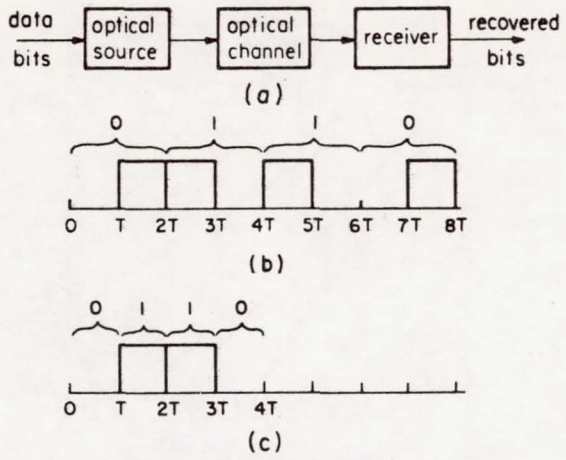


Fig. 1. (a) Digital optical system. (b) PPM energy waveform for 0110. (c) On–off energy waveform for 0110.

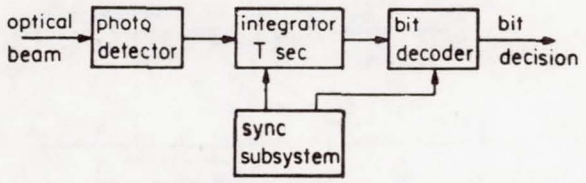


Fig. 2. Optical receiver for digital systems.

The digital receiver for the system is shown in Fig. 2. We shall assume that transmitter and receiver operate diffraction limited, so that the transmitted energy corresponds to optical energy in a single spatial mode of the optical beam. The received optical beam is photodetected, and its output is integrated over a T-second interval. The start–stop timing for this integration is provided by a synchronizing subsystem. In PPM, the bit decoder makes a comparison of the integrator output after the first T-second interval of each bit period with that after the second T-second interval, deciding a one or zero accordingly. In on–off keying a threshold test is made at the end of each bit time T, the bit decision depending upon whether the threshold is exceeded or not. The latter system requires accurate knowledge of the expected signal and noise energies in order to properly set the threshold, representing a serious disadvantage to on–off operation.

If the output of the photodetector is modeled [3] as a wide-band shot-noise process (detector bandwidth $\gg 1/T$), then the integrator output after T seconds of integration, beginning at time t, is proportional to the shot-noise counting process $\boldsymbol{k}(t, t + T)$, where

$$\boldsymbol{k}(t_1, t_2) = \text{number of photoelectrons in } (t_1, t_2). \quad (1)$$

In stating that the integrator value is proportional to (1), we have neglected additive circuit thermal noise, which implies the use of high-gain ideal photomultipliers in the photodetection operation. The counting process $\boldsymbol{k}(\cdot\ ,\ \cdot)$ of the photodetector shot noise is a random point process over the nonnegative integers. For the reception of an optical field over $(0, T)$, with the signal energy E and additive white Gaussian background noise of bandwidth B_0, the probability that the count value $\boldsymbol{k}(0, T)$ equals integer k is known to be [4]

$$\begin{aligned}\Pr[\boldsymbol{k}(0, T) = k] &\triangleq P_L(k; S, N, D) \\ &= \frac{(S)^k}{(1 + N_0)^{D+k+1}} \exp\left[-\frac{S}{1 + N_0}\right] L_k^D\left(-\frac{S}{(1 + N_0)N_0}\right)\end{aligned} \quad (2)$$

where $S = GE/hf$ is the average signal count over $(0, T)$; $N_0 = G[\exp(hf/kT_c - 1)]^{-1}$ is the average noise count per mode due to background at temperature T_c; h is Planck's constant; f is the laser frequency; $D = 2B_0T$, G is the photomultiplier gain; and $L_k^D(x)$ is the Laguerre polynomial in x of order D and index k:

$$L_k^D(x) = \sum_{i=0}^{k} \binom{k + D}{k - i} \frac{(-x)^i}{i}. \quad (3)$$

The parameter D is the count dimension or time–bandwidth product. Physically, $D + 1$ is the number of temporal modes observed during the T-second counting interval. The density $P_L(k; S, N, D)$ is called a Laguerre counting density and is exact for $D = 0$ and $D \gg 1$, but is only approximate for $D \approx 1$. (This is due to the fact that (2) requires equal eigenvalues in the expansion of the energy function, which is only approximately true for low values of D.) The received average signal energy E over the time interval T can also be written as $E = Q_sT$, where Q_s is the received average power. We then have, alternately,

$$S = (GQ_s/hf)T = \mu_sT \quad (4)$$

where μ_s is the average count per second (count rate) due to the signal.

Under typical operating conditions, we generally have $N_0 \ll 1$ and $D \gg 1$, and (2) asymptotically approaches the Poisson density [4]

$$\begin{aligned}\Pr[\boldsymbol{k}(0, T) = k] &\triangleq P_p(k; S + N) \\ &\equiv \frac{(S + N)^k}{k!} \exp[-(S + N)]\end{aligned} \quad (5)$$

where $N = DN_0$ represents the total noise count in all modes. (For visible wavelengths, N_0 is generally on the order of 10^{-7}–10^{-6} counts/mode. An optical system at 10 μ operating with a 1-A optical filter and $T = 10^{-6}$ s, will generate a D of about 400.) Note that with the Poisson assumption, the count probability depends only upon the sum of the signal and noise count. That is, the count statistics do not distinguish between the effect of signal energy or noise energy, but are determined solely by their cumulative energy.

II. Error Probabilities

If we transmit a binary PPM signal with fixed signal energy in the signaling interval, then the probability of making a bit error is simply the probability that the

count in the nonsignaling interval exceeds or equals that of the signaling interval. (If the counts are equal, an equally likely random choice is made concerning that particular bit.) If we denote k_i as the count in the ith interval, $i = 1, 2$, of a bit, then the average error probability PE is

$$\mathrm{PE} = \tfrac{1}{2}\Pr[k_2 > k_1 \mid \text{one sent}] + \tfrac{1}{2}\Pr[k_1 > k_2 \mid \text{zero sent}] + \tfrac{1}{2}\{[\tfrac{1}{2}\Pr k_2 = k_1 \mid \text{one sent}] + \tfrac{1}{2}[\Pr k_1 = k_2 \mid \text{zero sent}]\}. \quad (6)$$

From the symmetry of the transmission method, some of the terms above combine and the result simplifies. Thus, for Laguerre counting, (6) becomes

$$\mathrm{PE}_L(S, N_0, D) = \sum_{k_1=0}^{\infty}\sum_{k_2=k_1}^{\infty} \gamma_{k_1 k_2} P_L(k_1; S, N_0, D) P_L(k_2; 0, N_0, D) \quad (7)$$

where $\gamma_{k_1 k_2} = \frac{1}{2}$ for $k_2 = k_1$ and is one otherwise. (The $\frac{1}{2}$ factor accounts for the effect of equal interval counts.) Note that the error probability using Laguerre counting PE_L depends explicitly on the count dimension D (time–bandwidth product). If the Poisson assumption is applicable, the probabilities in (7) are replaced by those of (5), and we have

$$\mathrm{PE}_p(S, N) = \sum_{k_1=0}^{\infty}\sum_{k_2=k_1}^{\infty} \gamma_{k_1 k_2} P_p(k_1, S + N) P_p(k_2, N). \quad (8)$$

We see that the Poisson error probability PE_p depends only upon the parameter D through the total noise count $N = DN_0$. The Poisson error probability is easier to compute than that using Laguerre counting, and parametric studies of (8) have been extensively published [2], [5]. A typical plot of PE_p is shown in Fig. 3 as a function of the signal count S. Some PE_L points obtained by computing (7) at the same total noise level are superimposed. Further comparisons of Poisson and Laguerre error probabilities, in terms of the parameters involved, are discussed in [6]. The primary conclusion is that at low noise levels ($N \ll 1$), it can be conjectured that $\mathrm{PE}_L \approx \mathrm{PE}_p$ for moderate ($D \approx 100$) dimensions.

When on–off keying is used and a threshold test is made at the end of each pulse time T, an error is made whenever the integrator value is on the incorrect side of threshold. If K is the *a priori* selected threshold count value, then the error probability becomes

$$\mathrm{PE}_p(S, N) = \frac{1}{2}\sum_{k=0}^{K} \gamma_K P_p(k, S + N) + \frac{1}{2}\sum_{k=K}^{\infty} \gamma_k P_p(k, N) \quad (9)$$

where again $\gamma_K = \frac{1}{2}$ for $k = K$ and is one otherwise. For Laguerre statistics, the probabilities on the right should be replaced by the P_L terms in (2). For Poisson counting the sums in (9) are cumulative Poisson probabilities and are well tabulated (e.g., see [9]).

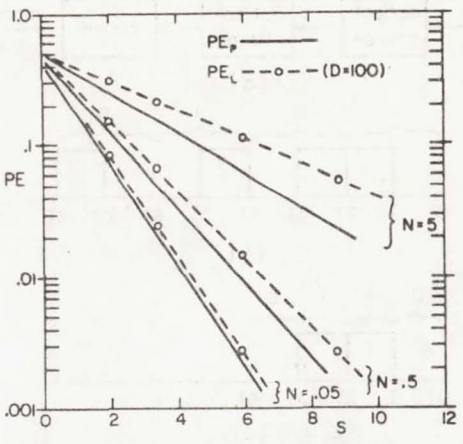


Fig. 3. Error probabilities versus signal count for PPM. N—noise count; D—count dimension.

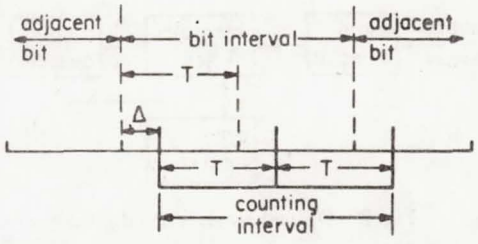


Fig. 4. Effect of timing error on counting interval in PPM.

III. Timing-Error Effects in PPM

The primary assumption in (8) is that the bit timing is perfect and the decoder counts photoelectrons exactly over the two T-second intervals that constitute a bit. If a time offset of Δ seconds occurs during a bit period, due to timing errors in synchronization lockup, then the counting occurs over an offset interval. That is, the decoder starts and stops counting over a T-second interval that is displaced by Δ seconds from that containing the bit information, as shown in Fig. 4. As a result only a portion of the true signal energy is included in the signal count, while some signal energy may contribute to the count in the adjacent interval, causing intersymbol interference in the form of energy spillover. The effect of this interference depends upon the form of the adjacent bit; i.e., whether it contains signal energy or not. Assuming a positive timing offset ($0 < \Delta < T$), the various effects on the counting statistics are summarized in Table I, where μ_s is the average signal count rate in (4). If we let $S = \mu_s T$ be the average count over T due to signal energy, and assume equiprobable bits, the error probability for a positive timing error Δ, averaging over all possibilities given in Table I, is then

$$\begin{aligned}\mathrm{PE}_p \mid \Delta = {} & \frac{1}{2}\sum_{k_1=0}^{\infty}\sum_{k_2=k_1}^{\infty} \gamma_{k_1 k_2} P_p[k_1, S(1-\epsilon)+N] P_p[k_2, N+S\epsilon] \\ & + \frac{1}{4}\sum_{k_1=0}^{\infty}\sum_{k_2=k_1}^{\infty} \gamma_{k_1 k_2} P_p[k_1, S(1-\epsilon)+N] P_p[k_2, N] \\ & + \frac{1}{4}\sum_{k_2=0}^{\infty}\sum_{k_1=k_2}^{\infty} \gamma_{k_1 k_2} P_p[k_1, S\epsilon+N] P_p[S+N] \end{aligned} \quad (10)$$

TABLE I

Transmitted Bit	Subsequent Bit	Pr $X(0, T) = k_1$	Pr $X(T, 2T) = k_2$
1	0	$P_p[k_1, \mu_s(T - \Delta) + N]$	$P_p[k_2, N]$
1	1	$P_p[k_1, \mu_s(T - \Delta) + N]$	$P_p[k_2, N + \mu_s\Delta]$
0	1	$P_p[k_1, \mu_s\Delta + N]$	$P_p[k_2, \mu_sT + N]$
0	0	$P_p[k_1, \mu_s\Delta + N]$	$P_p[k_2, \mu_s(T - \Delta) + N]$

where $\epsilon = \Delta/T$ is the percentage timing error. The error probability for negative time shifts will be identical to the preceding, when all possibilities are considered, if we interpret $\epsilon = |\Delta|/T$ when $\Delta < 0$. Note that if each of the double-sum terms in (10) is compared to (8), which assumed perfect timing, we can rewrite (10) as

$$\mathrm{PE}_p \mid \Delta = \tfrac{1}{2}\,\mathrm{PE}_p\,(S', N') + \tfrac{1}{4}\,\mathrm{PE}_p(S'', N) + \tfrac{1}{4}\,\mathrm{PE}_p\,(S'', N') \qquad (11)$$

where

$$S' = S(1 - 2\epsilon) \qquad (12a)$$

$$S'' = S(1 - \epsilon) \qquad (12b)$$

$$N' = N + S\epsilon. \qquad (12c)$$

Thus, timing errors in PPM can be accounted for by merely reinterpreting the effective signal and noise count per T interval while assuming perfect timing. Note that the timing errors always act to reduce the effective signal energy, while increasing the effective noise, the overall result degrading the error probability. It is important to realize that the fact that the spilled over signal energy appears as effective noise energy is intrinsic in the Poisson assumption and is valid as long as (8) describes the error probability.

A plot of (11), obtained by digital computation, is shown in Fig. 5 for positive or negative timing errors. The results show a relatively fast increase in PE (system degradation) as the offset $|\Delta|$ is increased. The system is essentially ruined (PE ≈ 0.5) when $\epsilon \approx 0.5$ or when $|\Delta| \approx T/2$. This is the point where the effective signal-to-noise ratios S'/N' and S''/N' are equal to or less than unity.

A lower bound to the system performance as $S \to \infty$ is included, obtained by envoking the fact that at low noise levels Poisson error probabilities and Laguerre error probabilities with the same total noise are roughly equal, as pointed out before. Since the PE_L monotonically increases with the parameter D, the use of PE_L at $D = 0$ will serve as a lower bound for error probability. When the signal has count S and the additive noise count is N, (7) with $D = 0$ is

$$\mathrm{PE}_L \mid_{D=0} = \left(\frac{1}{1+N}\right)^2\left(\frac{1+2N}{2}\right) \cdot \exp\left[\frac{-S}{1+N}\right] \sum_{k=0}^{\infty}\left(\frac{1}{1+N}\right)^{2k} L_k\left(\frac{S}{N(1+N)}\right). \qquad (13)$$

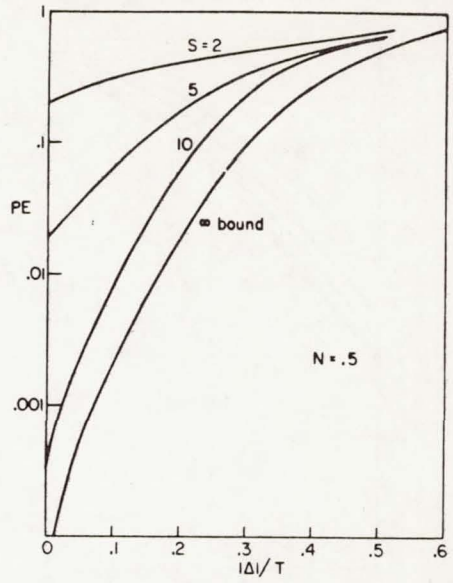


Fig. 5. Error probability versus timing error in PPM. S—signal count; N—noise count.

By applying a Laguerre identity [8, eq. 8.975] and manipulating algebraically, the above becomes

$$\mathrm{PE}_L \mid_{D=0} = \tfrac{1}{2}\exp\left[\frac{-S}{1+2N}\right]. \qquad (14)$$

If we substitute the effective S and N from (12) into (14), and use this as a lower bound for each term in (11), we have

$$\mathrm{PE}_p \mid \Delta \approx \mathrm{PE}_L \mid \Delta \geq \mathrm{PE}_L \mid \Delta;\ D = 0$$
$$= \tfrac{1}{4}\exp\left[\frac{-S(1-2\epsilon)}{1+2N+2\epsilon S}\right] + \tfrac{1}{8}\exp\left[\frac{-S(1-\epsilon)}{1+2N}\right] + \tfrac{1}{8}\exp\left[\frac{-S(1-\epsilon)}{1+2N+2\epsilon S}\right]. \qquad (15)$$

Now as $S \to \infty$,

$$\lim_{S\to\infty} \mathrm{PE}_p \mid \Delta \geq \tfrac{1}{4}\exp\left[-\frac{1-2\epsilon}{2\epsilon}\right] + \tfrac{1}{8}\exp\left[-\frac{(1-\epsilon)}{2\epsilon}\right]. \qquad (16)$$

The above lower bound depends only upon ϵ and is plotted as the $S = \infty$ curve in Fig. 5. The result is interesting in that it shows that even as $S \to \infty$, a relatively sharp system degradation can still be expected. This can be attributed again to the fact that timing errors cause a portion of the signal energy to appear as noise energy. Therefore, even though an "infinite" signal energy is available, there is consequently an "infinite" noise energy present, whenever $\epsilon \neq 0$, the overall result not dependent upon S at all as (16) illustrates.

The behavior of PE_p at different noise counts is shown in Fig. 6 for fixed value of S. Again, even with negligible background noise, the system degrades in a similar fashion with increasing timing error.

IV. Timing Error Effects With On–Off Keying

When on–off keyed data bits are transmitted and threshold tests are used for bit decisions at the decoder,

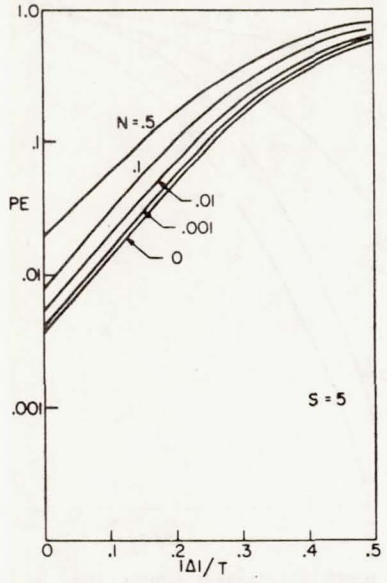


Fig. 6. Error probability versus timing error in PPM. S—signal count; N—noise count.

the effect of timing errors can be determined by a procedure similar to the PPM case. The actual bit decisions will be influenced by the adjacent bit (the subsequent bit when $\Delta > 0$, the former bit when $\Delta < 0$), just as in the previous case. If we consider the four possible combinations of transmitted and adjacent bits, and the associated error probability for each, the total error probability when a threshold K is used and an offset Δ occurs, is then

$$\mathrm{PE}_p \mid \Delta = \underbrace{\frac{1}{4}\sum_{k=0}^{k}\gamma_k P_p[S+N]}_{11} + \underbrace{\frac{1}{4}\sum_{k=K}^{\infty}\gamma_k P_p[N]}_{00}$$

$$+ \underbrace{\frac{1}{4}\sum_{k=0}^{k}\gamma_k P_p[S(1-\epsilon)+N]}_{10} + \underbrace{\frac{1}{4}\sum_{k=K}^{\infty}\gamma_k P_p[\epsilon S+N]}_{01} \quad (17)$$

where again $\epsilon = |\Delta|/T$ and S, N are the received signal and noise counts, respectively. The symbols below each sum represent the combination of data bits causing the corresponding error probability, with the left-hand bit the transmitted bit and the other the adjacent bit. Comparison of (17) with (9) allows us to write

$$\mathrm{PE}_p \mid \Delta = \tfrac{1}{2}\,\mathrm{PE}_p\,(S, N) + \tfrac{1}{2}\,\mathrm{PE}_p\,(S', N') \quad (18)$$

where the terms on the right are error probabilities with perfect timing, and S' and N' are defined in (12). We again observe that timing-error effects can be interpreted as degradations in signal energy and increases in noise energy in a perfectly timed system. Note that timing errors are exhibited only in the second term in (18), and can be attributed to the last two terms in (17), where the adjacent bit is opposite from the true bit. The error probabilities in (18) depend upon the choice of threshold K used for decisioning. For a given design value of S and N, the threshold K that minimizes (9) can be determined by differentiation, and shown to be

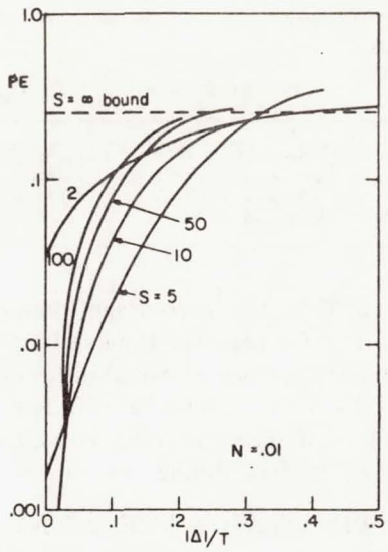


Fig. 7. Error probability versus timing error in on–off keying. S—signal count; N—noise count.

$$K = \frac{S}{\log\,(1 + S/N)}. \quad (19)$$

With this threshold, (18) is plotted in Fig. 7 as a function of timing offset for several values of S and N. The curves manifest similar behavior as in PPM, except the degradation is faster, and the curves exhibit crossovers. That is, at small offsets increasing S decreases error probability, but at larger offsets the opposite is true. An examination of the sums of (17) will reveal that for $N \ll 1$, $S \gg 1$ the first three terms tend to zero and the resulting $\mathrm{PE}_p \mid \Delta$ is directly attributable to the last term; i.e., the error probability when a zero is sent and the adjacent bit is a one. In the limits as $S \to \infty$, it follows that even though $K \to \infty$ [see (19)], this latter probability becomes exactly one for any $\epsilon \neq 0$. The overall $\mathrm{PE}_p \mid \Delta$ therefore becomes 0.25, and the result is plotted as the $S = \infty$ curve in Fig. 7. The behavior of all these curves can be directly attributed to the fact that optimal on–off keying requires proper threshold selection, and timing offsets cause changes in effective signal and noise energies and, hence, suboptimal operation. As these effective energies become widely different from the design energies, the resulting system performance is severely degraded.

V. Random Timing Errors

The timing error that does in fact occur during a bit interval depends upon the synchronization subsystem and its performance in maintaining time lock. This is generally accomplished by tracking a transmitted sync signal with a locally generated sync signal using a feedback tracking loop for error control. The timing error Δ is therefore the tracking error between the received and locally generated sync signals, and in reality should be considered as a random process in t. In typical operation, however, the loop-tracking bandwidth is much less than

the bit frequency $1/T$, and the assumption of a constant timing error during a given bit interval is essentially valid. The error is however random, and its statistics will depend upon the tracking-loop model. When sinusoidal sync signal at the bit frequency $1/T$ are used, and tracking is accomplished by a phase-lock loop (PLL) following photodetection, the steady-state probability density of Δ is given by

$$p(\Delta) = \frac{2\pi}{T} p_\varphi\left(\frac{2\pi\Delta}{T}\right) \tag{20}$$

where $p_\varphi(\varphi)$ is the density of the loop-tracking phase error φ. This latter density has been investigated for a system using a separate optical channel (different optical frequency) for transmitting the sync information [7]. When the sync channel is in quantum-limited operation (high-gain photomultiplication and negligible background energy), the steady-state probability density of the phase error has been approximated by computer solution of the Smoluchowski–Kolmogorov equation [10]. The latter equation is a nonlinear partial differential equation for the probability density of an output random process (in this case, the tracking error) of a dynamical system (the phase-tracking loop) when forced by an input random process (the photodetector output of the sync channel). The results of this computer solution, reported in [7], are shown in Fig. 8. The phase-error density depends only upon the parameter

$$\alpha = \frac{\mu_{sc}}{2B_L} \tag{21}$$

where B_L is the tracking-loop bandwidth and μ_{sc} is the average count rate due to the sync signal, the latter directly related to the received power in the sync channel. The parameter α is therefore the average number of sync-signal counts occurring in the time period $1/2B_L$. The bandwidth B_L must be selected large enough to allow suitable dynamical tracking of the incoming sync phase shifts (due to Doppler, range uncertainty, and oscillator phase jitter). For $\alpha \geq 3$, the phase densities are, to a good approximation, given by

$$p_\varphi(\varphi) = \frac{\exp[\alpha \cos\varphi]}{2\pi I_0(\alpha)}, \qquad |\varphi| < \pi \tag{22}$$

where $I_0(\alpha)$ is the imaginary Bessel function. Equation (22) may be recognized as the steady-state density associated with tracking a sync tone in the presence of additive Gaussian noise [11]. This implies that the optical sync channel performs identical to the microwave sync channel for reasonably high ($\alpha \geq 3$) count rates.

An average timing-error probability $\overline{\text{PE}}$ can be computed by averaging the PE | Δ in Figs. 5 and 7 over the random timing errors, using the density $p(\Delta)$ obtained from (20). That is,

$$\overline{\text{PE}} = \int_{-\infty}^{\infty} [\text{PE} \mid \Delta] p(\Delta)\, d\Delta. \tag{23}$$

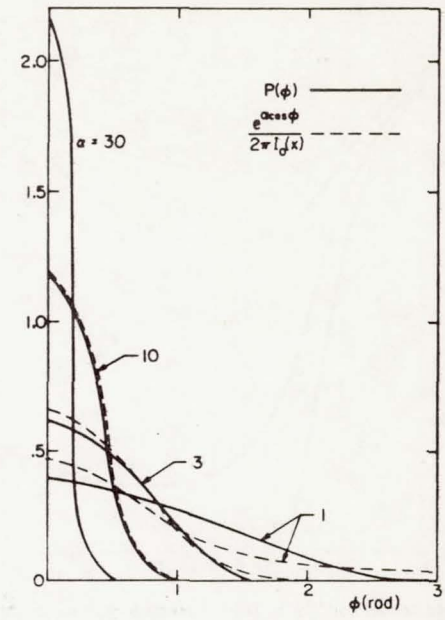


Fig. 8. Probability densities of tracking errors.

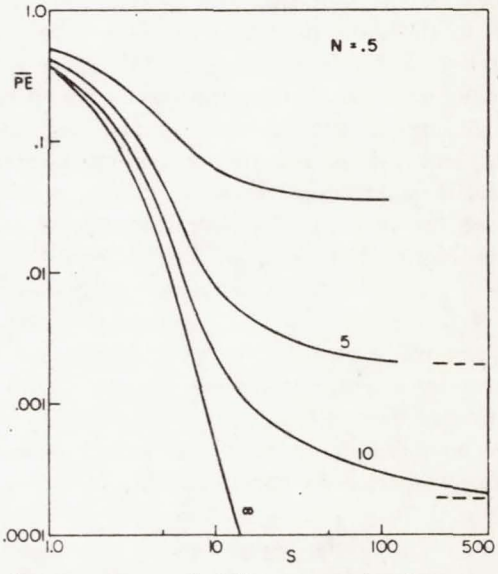


Fig. 9. Average error probability versus signal count for PPM. N—noise count; α—sync signal count.

The integral in (23) was evaluated using a point-by-point integration of the densities in Fig. 8. The results are plotted in Fig. 9 for PPM and in Fig. 10 for on–off keying, showing $\overline{\text{PE}}$ as a function of transmitted data signal count S, for a fixed background noise count N and several signal counts α in the sync channel. The results indicate the average effect of imperfect timing, exhibiting the usual falloff in error probability with increasing signal energy, followed by a flattening (Fig. 9) and bottoming (Fig. 10) of performance as S is increased. The values of the minimum $\overline{\text{PE}}$ depends upon the tracking-loop signal count. In PPM the minimum asymptotes plotted in Fig. 9 are those obtained by averaging the $S = \infty$ curve in Fig. 5 over the densities in Fig. 8. In on–off keying $\overline{\text{PE}}$ actually begins increasing after achieving a minimum value, even though S continues to increase.

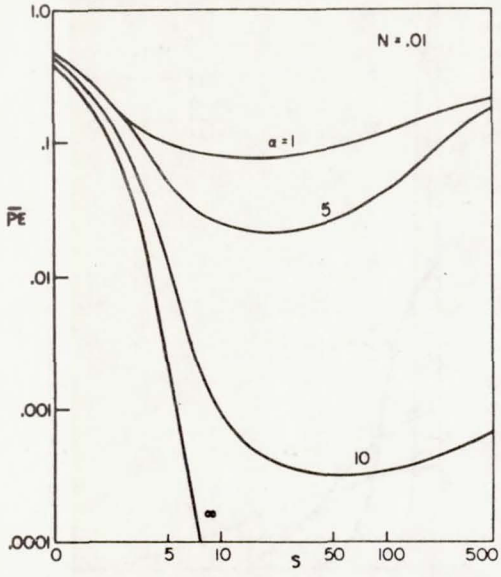


Fig. 10. Average error probability versus signal count for on–off keying. N—noise count; α—sync signal count.

This is due to the fact that the system is more "mismatched" in threshold design at the higher values of S. This latter fact tends to favor PPM operation over on–off keying when combating imperfectly timed systems. This bottoming of $\overline{PE}$ in both systems is extremely important since it represents a residual nonreducible error probability that depends only upon the sync system, and cannot be overcome by increasing the bit energy to the data signal. For example, we see from Fig. 9 that with $\alpha = 5$ and $N = 0.5$ we can never achieve an error probability less than 2×10^{-3}, no matter how much pulse energy we transmit. To determine these residual values for other design parameters, the curves for $PE \mid \Delta$ must be first generated then averaged as in (23).

It may be pointed out that the same residual effect due to imperfect tracking occurs in the additive Gaussian noise channel (microwave system instead of optical) when using phase-shift keyed binary transmission. In this latter case, as data signal energy becomes infinite, $PE \to 0$ as long as the tracking error is less than $\pi/2$ rad, and $PE \to 1$ for $\varphi \geq \pi/2$. Thus the residual error probability is simply the probability that the loop error exceeds $\pi/2$. In Fig. 5 we see that as $S \to \infty$ we do not obtain zero PE (except at $\epsilon = 0$), and the residual $\overline{PE}$ tend to be higher than the comparable microwave case. That is, to obtain the same residual $\overline{PE}$, the optical system requires more sync power.

References

[1] *Proc. IEEE* (Special Issue on Optical Communication), vol. 58, pp. 1407–1786, Oct. 1970.
[2] R. Gagliardi and S. Karp, "M-ary Poisson detection and optical communications," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 208–216, Apr. 1969.
[3] S. Karp, E. O'Neill, and R. Gagliardi, "Communication theory for the free-space optical channel," *Proc. IEEE*, vol. 58, pp. 1611–1626, Oct. 1970.
[4] S. Karp and J. R. Clark, "Photon counting: A problem in classical noise theory," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 672–680, Nov. 1970.
[5] W. K. Pratt, *Laser Communications*. New York: Wiley, 1969, ch. 9.
[6] R. Gagliardi, "Photon counting and Laguerre detection," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-18, pp. 208–211, Jan. 1972.
[7] M. Haney and R. Gagliardi, "Optical synchronization–Phase locking with shot noise processes," USCEE Rep. 396, Aug. 1970.
[8] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series, and Products*. New York: Academic Press, 1965.
[9] M. Abramowitz and I. Stegun, Eds. *Handbook of Mathematical Functions* (Applied Mathematics Series 55). Washington, D. C.: NBS, 1964, Table 26.
[10] D. Middleton, *Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960, ch. 10.
[11] A. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966, pp. 86–96.

Robert M. Gagliardi (S'57–M'61) was born in [redacted], on [redacted]. He received the B.S. degree in electrical engineering from the University of Connecticut, Storrs, in 1956, and the M.S. and Ph.D. degrees in engineering from Yale University, New Haven, Conn., in 1957 and 1960, respectively.

From 1958 to 1960 he was an Instructor at the New Haven Engineering College. In 1960 he joined the Information Study Section, Space System Divison, Hughes Aircraft Company, Culver City, Calif., where he was involved in problems in telemetry and communication systems. He is presently an Associate Professor in the Department of Electrical Engineering, University of Southern California, Los Angeles, and a Consultant to the Hughes Aircraft Company.

Dr. Gagliardi is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

Reprinted by permission from
IEEE TRANSACTIONS ON COMMUNICATIONS
Vol. COM-22, No. 10, October 1974

Synchronization Using Pulse Edge Tracking in Optical Pulse-Position Modulated Communication Systems

R. M. GAGLIARDI, MEMBER, IEEE

***Abstract*—A pulse-position modulated (PPM) optical communication system using narrow pulses of light for data transmission requires accurate time synchronization between transmitter and receiver. The presence of signal energy in the form of optical pulses suggests the use of a pulse edge-tracking method of maintaining the necessary timing. In this report the edge-tracking operation in a binary PPM system is examined, taking into account the quantum nature of the optical transmissions. Consideration is given first to "pure" synchronization using a periodic pulsed intensity, then extended to the case where position modulation is present and auxiliary bit decisioning is needed to aid the tracking operation. Performance analysis is made in terms of timing error and its associated statistics. Timing error variances are shown as a function of system signal-to-noise ratio.**

I. INTRODUCTION

The successful operation of any digital communication system requires accurate time synchronization between the transmitter and receiver. In optical digital systems a common procedure is to use a noncoherent pulse-position modulation (PPM) mode of operation using narrow pulses of light intensity to carry the data [1]. The presence of signal energy in the form of optical pulses suggests the use of a pulse edge-tracking method of maintaining the necessary time synchronization. In pulse edge tracking the edges of the transmitted pulses are used as timing markers to adjust the synchronization of the receiver. When the optical pulses are transmitted as a periodic pulse train of known fixed frequency, the edge tracking corresponds to "pure" synchronization, in that the transmitted edges always occur at periodic points in time. When position modulation is present, however, the pulses of light are shifted according to the data, and the edge-tracking operation must be modified in order to maintain receiver timing. The latter type of synchronization is often called modulation-derived synchronization, or "impure" syncing, since the timing must be derived from, or accomplished in the presence of, the data modulation. In this paper we examine the pure and impure edge-tracking operation in an optical binary PPM system, taking into account the quantum nature of the light transmission. Performance comparisons are made in terms of the instantaneous timing error of the receiver and its associated statistics. The effect of imperfect timing on the overall data decoding operation has been studied elsewhere [2] and will not be considered here.

The time synchronization problem has of course received considerable attention in the past for the additive Gaussian noise channel, and the interested reader is referred to the presentations in recent books by Stiffler [3], Lindsey [4, ch. 3], and Lindsey and Simon [5]. Although the approach here parallels these earlier studies, the quantum nature of the optical channel produces equations significantly different than those of the purely Gaussian channel. Similar mathematical differences were previously observed with the optical

Paper approved by the Associate Editor for Communication Theory of the IEEE Communications Society for publication without oral presentation. This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-108-104. This grant is part of the research program at NASA's Goddard Space Flight Center, Greenbelt, Md. Manuscript received November 27, 1972; revised June 5, 1974.

The author is with the Department of Electrical Engineering, University of Southern California, Los Angeles, Calif. 90007.

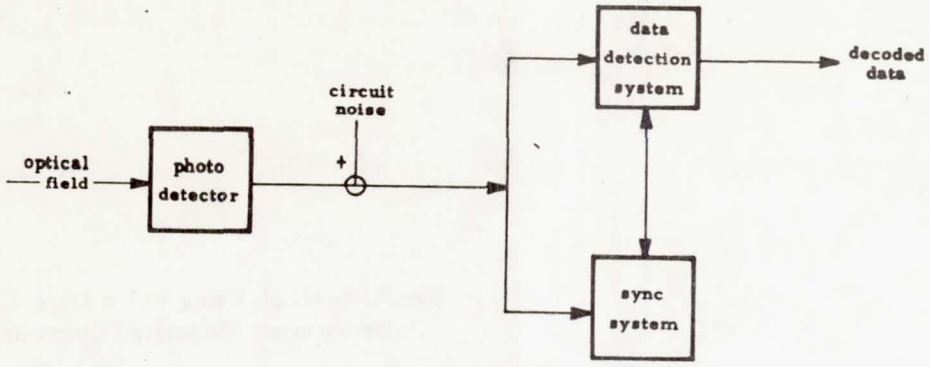


Fig. 1. An optical digital PPM receiver.

channel when considering detection [6], waveform estimation [7], and sinusoidal phase tracking [8]. The results here for pulse synchronization are reminiscent of (but not identical to) the effects of tracking in an additive shot-noise environment [9].

II. SYSTEM DESCRIPTION

A block diagram of an optical digital system is shown in Fig. 1. In normal operation the incident optical field is photodetected, and the recovered signal is processed in both a data detection channel and in a synchronization channel, the latter providing the timing for the former. In this paper only the synchronization subsystem will be considered. In a PPM noncoherent mode of operation, digital information is transmitted by position modulating a pulse of light intensity during each digital word interval. Thus, in a binary system, the light energy is transmitted in one of two adjacent bit subintervals, representing a binary one or binary zero, as shown in Fig. 2(a). Detection in the data channel is made by photoelectron counting (physically, short-term integration of the photodetector output, which is equivalent to energy detection of the optical field) during each subinterval, deciding on the position with the highest count as containing the transmitted bit. Timing for the starting and stopping of each counting interval is provided by the synchronization subsystem, and timing errors (offsets between received and integrated bit intervals) lead directly to system degradation. Continual timing information is necessary in order to maintain bit timing in spite of the time delay variations that may occur during optical transmission.

The system of Fig. 1 can operate with one of two different synchronization formats. For pure sync operation, an unmodulated sync signal (herein considered as a periodic train of optical pulses at the bit-rate frequency), shown in Fig. 2(b), is transmitted intermittantly in place of the data to allow receiver lockup, and the resulting timing is used to decode the subsequent data transmissions until the system is retimed with the next sync burst. In impure sync generation, the PPM data are transmitted continuously and the timing is extracted from the data. The first procedure allows pure synchronization but must sacrifice data during the timing operation. The second method allows uninterrupted data transmission, and is obviously the preferred method of operation, but requires modulation-derived synchronization. For this reason considerable interest exists in developing the latter system and in determining its achievable performance.

In pulsed optical systems both the pure and impure sync systems can employ edge tracking for timing. An edge-tracking subsystem makes use of the fact that a pulse edge always occurs at the center of each bit interval (see Fig. 2) and can therefore be used to sync an identical pulse train at the receiver. The subsystem, to be described in Sections III and IV, employs a feedback loop to essentially measure timing errors between the received and receiver pulse edges, using the error to correct the latter signal. In pure synchronization the pulse edge at the center always corresponds to the trailing edge of a pulse. When PPM is present, however, the edge may represent either a leading or trailing edge, and this polarity must be determined for successful loop operation. To accomplish this, the sync subsystem in the modulated case employs an auxiliary decision-making loop that operates in conjunction with the edge-tracking loop (see Fig. 7). This auxiliary loop essentially decides which type of edge (i.e., which data bit) is being received, using the decision to augment a delayed version of the standard edge-tracking loop. Similar systems have been previously proposed [10].

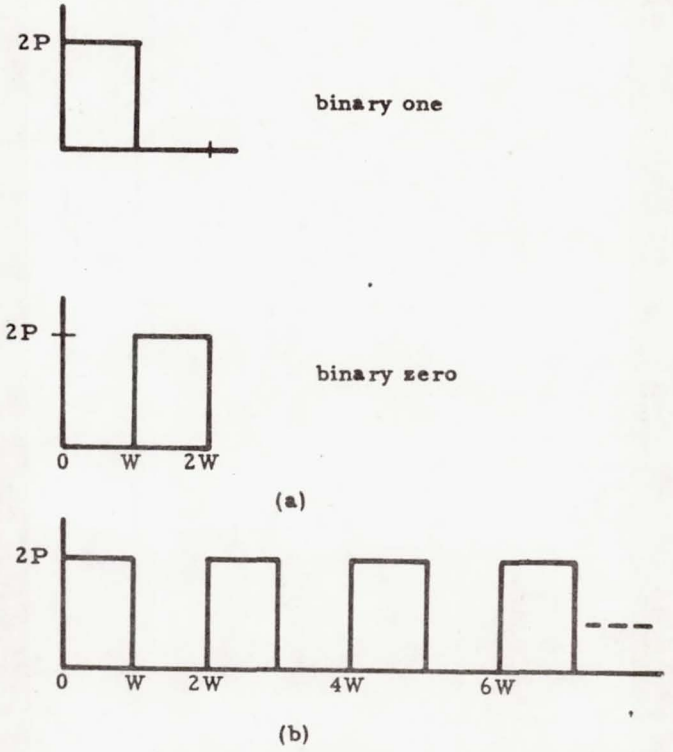


Fig. 2. Intensity waveforms. (a) PPM bit intensities. (b) Pure sync intensity.

Continual or updated timing is necessary to overcome the unintentional variations in transmission delay, due to Doppler, relative receiver motion, etc. If the basic assumption is made that these delay variations are slow relative to the optical pulsewidth, then their only effect is to vary the time location of the optical pulse without distorting its shape. Thus, if $I(t)$ represents the light intensity at the receiver with no delay variations, and if τ_1 is the time-varying delay occurring, then the recovered field intensity is given by $I(t - \tau_1)$. Here it is tacitly implied that τ_1 is a function of t which changes slowly with respect to the pulsewidth of $I(t)$. Note, this latter condition is equivalent to the assumption that the bandwidth of τ_1 is much smaller than the bandwidth of $I(t)$. The principle objective of the edge-tracking loop is therefore to "track out" the unintentional time variations of τ_1 generated during the transmission of the optical field.

III. EDGE TRACKING OF A PERIODIC PULSE TRAIN (PURE SYNC)

In this section we first examine the edge-tracking operation when the received intensity corresponds to a periodic pulse train of light. This would represent the situation during pure synchronization operation when the data modulation is not present [actually, a periodic pulse train at the bit frequency can be considered as a continuous sequence of the binary one symbol in Fig. 2(a)]. The received field intensity with no delay variations is therefore given by

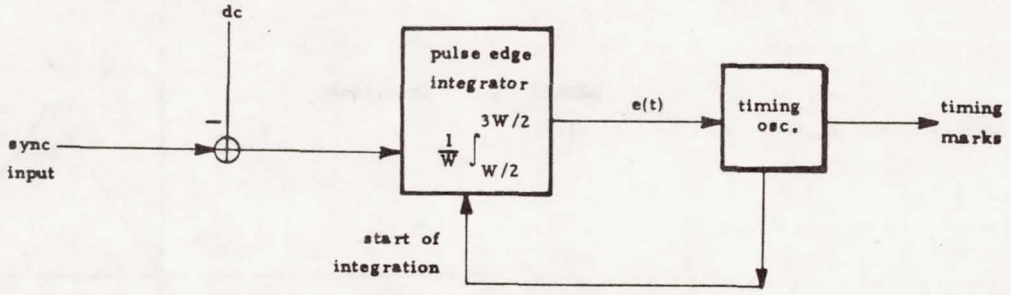


Fig. 3. Pulse edge-tracking subsystem.

$$I(t) = P[1 + p(t)] \tag{1}$$

where P is the average received field power per unit area, and $p(t)$ is the effective intensity modulation

$$p(t) = \begin{cases} 1 & 0 \le t \le W \\ -1 & W \le t \le T. \end{cases} \tag{2}$$

Here W is the pulsewidth and $T = 2W$ is the bit period. The above intensity is assumed to be received with the delay variations τ_1, which is equivalent to replacing $p(t)$ by $p(t - \tau_1)$.

The output of the photodetector in Fig. 1, operating over a single spatial mode of the optical field, is known to be the shot-noise current process [11]

$$i(t) = Ge \sum_{m=1}^{N(0,t)} \delta(t - t_m) \tag{3}$$

where $\delta(t)$ is the detector impulse function, e is the electron charge, G is the photomultiplication gain, $\{t_m\}$ are the random event times, and $N(0,t)$ is the electron Poisson counting process [i.e., $N(0,t)$ is the number of electrons occurring in the interval $(0,t)$]. The counting process has its average value related to the received field intensity $I(t)$ by

$$\text{avg } N(0,t) = \beta \int_0^t (I(t) + k_n)\, dt \tag{4}$$

where $\beta = \eta A/hf$, A is the detector receiving area, η is the receiver efficiency parameter, h is Planck's constant, f is the optical frequency, and βk_n is the average rate of arrival of background noise photons over the detector area. The photodetector output $i(t)$ of (3) will have added to it a white Gaussian circuit noise current $i_n(t)$, and the resulting signal, $x(t) \triangleq i(t) + i_n(t)$, provides the input to the synchronization edge-tracking subsystem, shown in detail in Fig. 3. A pulse-edge integrator is time controlled by a receiver timing oscillator, generating an error voltage used to readjust the oscillator. The latter, in addition, provides the timing markers for the data channel. The pulse-edge integrator consists simply of a W-seconds integration offset over the trailing edge of the received pulse. If the input to the integrator was the pulse train of Fig. 2(b) with zero average value (i.e., with its dc value removed) the offset integration occurs over portions of positive and negative values. If this latter integration had been timed to begin exactly halfway through the positive pulse (at $t = W/2$), the resulting integrated error value would be zero, no oscillator correction is necessary, and the system is in time sync. If a time difference occurred between start of integration and pulse half-interval point, a proportional error signal would be generated whose polarity depended on the direction of the time difference. This error voltage can be used to adjust the loop timing oscillator. The input dc removal can be accomplished easily by capacitor-coupled circuits, but is represented as a subtraction in the mathematical model of Fig. 3. Unfortunately, in the optical system of Fig. 1, the input to the loop is not a clean pulse sync train, but rather the shot-noise process of (3), containing the optically pulsed intensity of (1). In addition, this shot noise has added to it the additive circuit white-noise current $i_n(t)$. Hence, the error signal generated after the short-term loop integration is

$$\begin{aligned} e(t) &= \frac{1}{W}\int_{\tau_l}^{\tau_l+W} [i(t) + i_n(t) - Ge\beta(P + k_n)]\, dt \\ &= \frac{1}{W}\int_{\tau_l}^{\tau_l+W} i(t)\, dt + \frac{1}{W}\int_{\tau_l}^{\tau_l+W} i_n(t)\, dt - Ge\beta(P + k_n) \end{aligned} \tag{5}$$

where τ_l represents the start of the loop integration; i.e., the timing of the loop. The subtraction in (5) represents the removal of the average intensity from the loop input. The dependence of the right-hand side on t is implicit in the parameter τ_l, which varies as the loop attempts to track out the variations τ_1. Although τ_l is actually a function of t it is treated as a constant when integrating over the pulsewidth W seconds long. The latter fact is simply a restatement of the fact that the bandwidth of τ_l, which is roughly the same as that of τ_1, is much less than the repetition frequency $1/W$. After substituting the input processes, (5) can be rewritten as

$$e(t) = \frac{Ge}{W}[N(\tau_l, \tau_l + W)] + n(\tau_l) - Ge\beta(P + k_n) \tag{6}$$

where $n(\tau_l)$ is the Gaussian circuit noise random process, obtained by integrating $i_n(t)$ over $(\tau_l, \tau_l + W)$, and has zero mean and variance $N_0/2W$, with N_0 the one-sided circuit noise spectral level.

The performance of the tracker can be directly related to the instantaneous timing error between the received and the oscillator signal. This timing error is defined by

$$\tau \triangleq \tau_1 - \tau_l \tag{7}$$

where all parameters are actually functions of t. Using straightforward analog loop analysis, and recalling that the oscillator phase depends on the integral of the voltage controlling it, the timing error τ in (7) satisfies the integral-differential equation

$$\frac{d\tau}{dt} = \frac{d\tau_1}{dt} - Ke(t), \tag{8}$$

where K is the total loop gain. Since the error signal $e(t)$ depends on both τ_1 and τ_l, the equation is in general nonlinear in τ. Clearly, the solution for $\tau(t)$ in (8) necessarily evolves as a stochastic process due to randomness of $e(t)$ in (6). This is true even if the additive circuit noise $i_n(t)$ is set equal to zero (i.e., only background interference) due to the randomness of the shot-noise process.

Although the probability densities of $\tau(t)$ will be of ultimate interest, the behavior of the instantaneous mean value of $\tau(t)$ can be derived from (8). If we statistically average both sides, interchanging differentiation and averaging on the left, we obtain the equation

$$\frac{d\bar{\tau}(t)}{dt} = \frac{d\bar{\tau}_1(t)}{dt} - K\overline{e(t)}, \tag{9}$$

where the overbar denotes statistical average. Since the additive noise variable in (6) has zero mean, the mean-error voltage is given by the mean shot-noise count. The latter is the integrated count intensity over the integration intervals, as denoted by (4). Hence,

$$\begin{aligned} \overline{e(t)} &= \mathcal{E}_\tau\left\{\frac{Ge\beta}{W}\int_{\tau_l}^{\tau_l+W} [I(t) + k_n]\, dt - Ge\beta(P + k_n)\right\} \\ &= \mathcal{E}_\tau\left\{\frac{Ge\beta}{W}\int_{\tau_l}^{\tau_l+W} [P[1 + p(t - \tau_1)] + k_n]\, dt - Ge\beta(P + k_n)\right\} \\ &= \mathcal{E}_\tau\left\{\frac{e\beta GP}{W}\int_{\tau_l}^{\tau_l+W} p(t - \tau_1)\, dt\right\} \end{aligned} \tag{10}$$

where $\mathcal{E}_\tau$ is the expectation operator over the random variable τ, and $p(t)$ is given in (2). The above integral can be rewritten in terms of a receiver correlation. Define the function $y(t)$ by

$$y(t) = \begin{cases} 1 & 0 \le t \le W \\ 0 & \text{elsewhere.} \end{cases} \tag{11}$$

Then (10) becomes

$$\overline{e(t)} = \mathcal{E}_\tau\{e\beta GPR_{yp}(\tau)\} \tag{12}$$

where

$$R_{yp}(\tau) = \frac{1}{W}\int_{-\infty}^{\infty} p(t-\tau_1)y(t-\tau_l)\,dt. \tag{13}$$

Hence, the mean of the error process can be related to the correlation of the periodic intensity modulation $p(t)$ with the time function $y(t)$. The latter function can therefore be considered as the receiver "timing" signal produced by the loop. The correlation function $R_{yp}(\tau)$ for the functions $p(t)$ in (2) and $y(t)$ in (11) is plotted in Fig. 4. This correlation function is the mean-error function of the tracking loop, and is often referred to as the loop "S curve."

Equation (9) therefore becomes

$$\frac{d\bar{\tau}(t)}{dt} = \frac{d\bar{\tau}_1(t)}{dt} - (e\beta KGP)\mathcal{E}_\tau\{R_{yp}(\tau)\}. \tag{14}$$

The above is the differential equation of the mean timing error variable in the tracking loop. If $\tau(t)$ is confined to the linear range of $R_{yp}(\tau)$ (i.e., if $\tau \approx 0$ and the loop is tracking well) then we can approximate $R_{yp}(\tau) \approx 2\tau/W$ and $\mathcal{E}_\tau\{R_{yp}(\tau)\} \approx \mathcal{E}_\tau[2\tau/W] = 2\bar{\tau}/W$. Equation (14) then becomes a *linear* differential equation in terms of the mean-error process $\bar{\tau}(t)$. Furthermore, this linear equation corresponds to that of the linear feedback system in Fig. 5. The latter is often called the linear mean equivalent loop to Fig. 3, and is useful for analyzing or synthesizing based upon the mean timing error process. Note that in this equivalent system, the input delay variation τ_1 appears as the loop input, and the loop timing oscillator becomes a feedback loop integrator whose output is the timing process $\tau_l(t)$. The linear equivalent loop has a loop gain of $2eGK\beta P/W$ and a loop bandwidth[1] of

$$B_L = \frac{eGK\beta P}{2W}. \tag{15}$$

Note that the loop bandwidth depends directly on the received optical power P, which therefore appears as a parameter of the equivalent system. The loop bandwidth must be sufficiently wide (i.e., there must be sufficient loop gain) to track the expected time variations in τ_1.

Although mean-error performance in tracking the received delay variations can be determined from the linear mean system, the adverse effects due to the random nature of the optical field and circuit noise cannot be derived (note that the linear system is noiseless). In this case, the dynamical equation of the true system, (8), must be examined in detail for a complete statistical analysis. Unfortunately, the discrete nature of the timing error equation indicates that the statistics of the solution $\tau(t)$ will be highly nonstationary as the process evolves in time. An indication of the statistical properties of $\tau(t)$ can be obtained by examining the steady-state probability density of τ. This latter density, $f(\tau)$, is known to satisfy the Kolmogorov–Smoluchowski steady-state equation [4, Ch. 7]:

$$\sum_{j=1}^{\infty} \frac{(-1)^j}{j!}\frac{\partial^{j-1}}{(\partial\tau)^{j-1}}[K_j(\tau)f(\tau)] = 0, \tag{16}$$

where

$$K_j(\tau) = \lim_{\Delta t \to 0} \frac{\overline{(\Delta\tau)^j}}{\Delta t} \tag{17}$$

$$\Delta\tau = [\tau(t+\Delta t) - \tau(t)]$$

with suitable initial conditions and with the condition that $f(\tau)$ integrate to one. When the coefficients $K_j(\tau)$ exist, this equation provides a relation that must be satisfied by the steady-state density of the process $\tau(t)$. The equation is a partial differential equation

[1] In a linear feedback system, if $H(s)$ is the transfer function from loop input to feedback signal, then the loop bandwidth is defined by $B_L = \int |H(j\omega)|^2 d\omega/2\pi$. It essentially represents the bandwidth that the loop exhibits to the input.

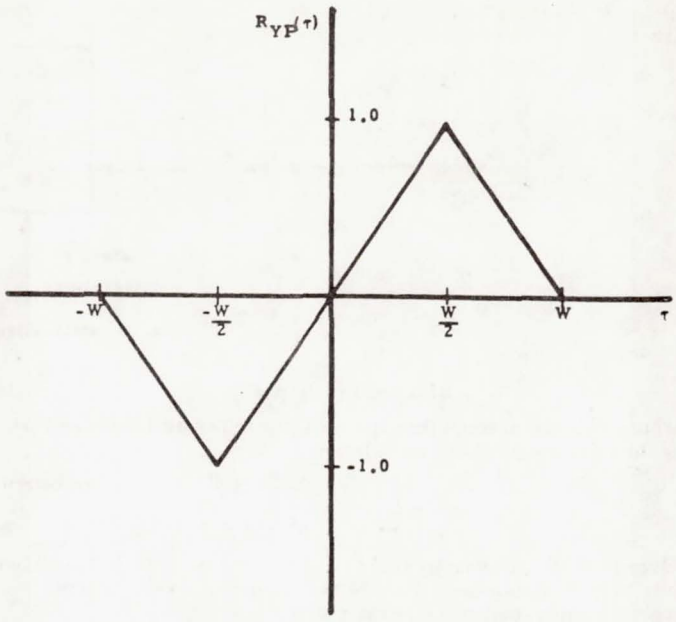


Fig. 4. Tracking error characteristic for pure sync.

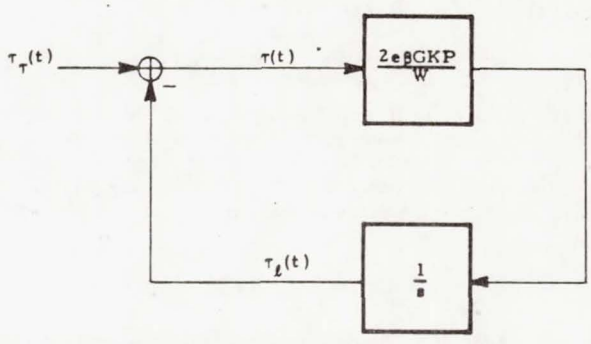


Fig. 5. Linear equivalent edge-tracking loop.

with variable coefficients and, in general, involves all orders of derivatives. The principle usefulness of (16), however, occurs when only the first few coefficients are nonzero. In particular, if $K_j(\tau) = 0$, $j \ge 3$, the resulting equation is the steady-state Fokker–Planck equation, and has been extensively studied [4, Ch. 7]. A Fokker–Planck equation implies a "continuous" process; i.e., processes that do not change significantly over a short-time period, while the more general equation of (16) would be associated with processes containing statistical jumps.

The calculation of the sequence of coefficients $K_j(\tau)$ requires determination of the moments of the error increment $\Delta\tau$ in (17). Consider again the system of (8) when tracking an intensity pulse having a constant-time shift $\tau_1(t) = \tau_0$. The timing error $\tau(t)$ therefore satisfies (8) with $d\tau_1/dt = 0$. The timing error variation Δt is then

$$\Delta\tau = \int_t^{t+\Delta t}\left(\frac{d\tau}{dz}\right)dz$$

$$= -K\int_t^{t+\Delta t} e(z)\,dz. \tag{18}$$

The coefficients in (17) can now be determined by using (6) in (18). Unfortunately, an exact calculation of the coefficients is hampered by the sampled data (short-term integrated) nature of the loop. (The integrator smooths the error signal, and produces in effect a second-order loop.) We shall consider instead a continuous first-order loop in which the integrator is neglected, and the resulting time averaged (over a period W) coefficients are used as an approximation to the desired steady-state coefficients. This is equivalent to assuming that the integration over the pulsewidth is extremely short relative to the time variations of the input process, and can be neglected. The subsequent time averaging of the coefficients is

similar to the smoothing produced by the loop. It is shown in Appendix A that the coefficients computed in this way become

$$K_j(\tau) = \begin{cases} -(Gek\beta P)R_{yp}(\tau), & j = 1 \\ (Gek)^2[\beta R_{y^2I}(\tau) + N_0/2(Ge)^2], & j = 2 \\ (-Gek)^j\beta[R_{y^jI}(\tau)], & j \geq 3 \end{cases} \quad (19)$$

where

$$R_{y^jI}(\tau) = \frac{1}{W}\int_{-\infty}^{\infty} y^j(t-\tau_l)[I(t-\tau_l) + k_n]\,dt. \quad (20)$$

For the loop signal $y(t)$ in (11), the above can be further simplified by noting that for all j

$$R_{y^jI}(\tau) = R_{yI}(\tau). \quad (21)$$

Thus, the steady-state Kolmogorov equation becomes

$$0 = (eGK\beta P)[R_{yp}(\tau)f(\tau)] + \frac{(eGK)^2}{2}\frac{d}{d\tau} \cdot \{[\beta R_{yI}(\tau) + N_0/2(Ge)^2]f(\tau)\} + \sum_{j=3}^{\infty}\frac{(eGK)^j}{j!}\frac{d^{j-1}}{d\tau^{j-1}}\{\beta R_{yI}(\tau)f(\tau)\}. \quad (22)$$

The infinite number of derivatives manifests the discontinuities of the error process caused by the quantum nature of the detected optical process. It is the form of this equation vis-a-vis the Fokker–Planck equation that theoretically separates optical tracking from tracking in additive Gaussian noise. This complication was previously noted by Ohlson [9] when dealing with added shot noise.[3] Although an exact solution for $f(\tau)$ is somewhat ambitious, some meaningful information and approximating solutions can be derived. In particular, consider the case where the system operates in near-lock operation, so that it may be assumed that $\tau \approx 0$. The instantaneous tracking error can therefore be considered to be confined to the linear range of $R_{yp}(\tau)$ and

$$R_{yp}(\tau) \approx 2\tau/W$$
$$R_{yI}(\tau) \approx (P + k_n). \quad (23)$$

Substituting into (22) and dividing by the coefficient of the second term yields the modified equation

$$0 = \alpha\tau f(\tau) + \frac{d}{d\tau}f(\tau) + \sum_{j=3}^{\infty} A_j\frac{d^{j-1}}{d\tau^{j-1}}f(\tau) \quad (24)$$

where

$$\alpha = \frac{2\beta P}{(WeGK/2)[\beta(P + k_n) + N_0/2(Ge)^2]} \quad (25)$$

$$A_j = \frac{2}{j!}\left(\frac{4}{W\alpha}\right)^{j-2}\left[\frac{\beta(P + k_n)(\beta P)^{j-2}}{[\beta(P + k_n) + N_0/(Ge)^2 2]^{j-1}}\right]. \quad (26)$$

Note that the coefficients A_j vary as $1/\alpha^{j-2}$, exhibiting a decreasing importance of the higher derivatives as the parameter α in (25) is increased. (The bracketed term in A_j is bounded by one and approaches one as the system approaches quantum limited operation, i.e., when $\beta P \gg \beta k_n + N_0/(Ge)^2 2$.) A physical interpretation to α can be introduced by noting the linear mean-equivalent loop of Fig. 5. Since it is often desirous to operate the tracking loop with a given loop bandwidth B_L the loop gain K is generally adjusted so as to obtain this value in (15). Thus, if B_L is the desired bandwidth, then we set $K = 2B_LW/Ge\beta P$, and (25) becomes

$$\alpha = \frac{2}{W^2}\left\{\frac{(\beta P)^2}{B_L[\beta(P + k_n) + N_0/2(Ge)^2]}\right\}. \quad (27)$$

When written in this way, the numerator in the braces is the square of the mean intensity of the received signal, while the denominator is effectively the total noise power occurring in the bandwidth B_L (since the level $\beta(P + k_n)$ is the two-sided shot-noise spectral level). Thus, α is proportional to a signal-to-noise power ratio, and the coefficients in A_j in (26) vary as an inverse power of this ratio. We therefore expect that solutions for $f(\tau)$ can be suitably approximated by solving truncated versions of (24) with fewer and fewer terms, as we increase the optical signal-to-noise ratio in (27).

Further properties of the solution for $f(\tau)$ for the in-lock operation case can also be derived. Transforming both sides of (24) indicates that the characteristic function of $f(\tau)$, $\varphi(\omega)$, satisfies the differential equation

$$-\alpha\frac{d\varphi(\omega)}{j\,d\omega} + j\omega\varphi(\omega) + \sum_{i=2}^{\infty} A_{i+1}(j\omega)^i\varphi(\omega) = 0. \quad (28)$$

This means

$$\ln\varphi(\omega) = -\frac{\omega^2}{2\alpha} + \frac{1}{\alpha}\sum_{i=2}^{\infty}\frac{A_{i+1}(j\omega)^{i+1}}{i+1} + C \quad (29)$$

where C is chosen to satisfy the unit area constraint on $f(\tau)$. Since the right-hand side is in the form of a power series in $j\omega$, the semi-invariants of the steady-state density can immediately be identified. Note that the first semi-invariant (mean value) is zero, the second semi-invariant (variance) is $1/\alpha$, and the higher semi-invariants are related to the $\{A_i\}$. Thus, the actual form of the solution density depends on these coefficients. As a limiting case, however, we see that if $\alpha \to \infty$, implying $A_i/\alpha \to 0$ for all $i \geq 3$, then $\ln\varphi(\omega) = -\omega^2/2\alpha$, corresponding to a zero mean, Gaussian density for τ, having variance $1/\alpha$. On the other hand, for $\alpha < \infty$ the higher coefficients can no longer be neglected, and the complete series in (29) must be included. Of course, as α is decreased in value, the increasing variance will cause the loop error to exceed the linear range of $R_{yp}(\tau)$, violating our assumption that the loop is in fact completely linear. Nonetheless it is important to recognize that even though the loop error density varies in form from the asymptotic Gaussian density for $\alpha \to \infty$ to the more complicated density defined by (29) as $\alpha \to 0$, the variance of the density for the linear system is always $1/\alpha$ [i.e., the second semi-invariant in (29)]. Assuming shot-noise limited operation $[\beta(P + k_n) \gg N_0/(Ge)^2 2]$, the loop error variance for the normalized delay variable $x = \tau/W$ is therefore

$$\sigma_x^2 = \frac{1/\alpha}{W^2} = \frac{1}{\Gamma}\left[1 + \frac{k_n}{P}\right] = \frac{1}{\Gamma}\left[1 + \frac{2\beta k_n/B_L}{\Gamma}\right] \quad (30)$$

where

$$\Gamma = 2\beta P/B_L. \quad (31)$$

The above is plotted in Fig. 6 as a function of the parameter Γ for several values of normalized noise energy $\beta k_n/B_L$. The curves, in essence, summarize the performance of a pure sync system operating with optical power P watts in a tracking bandwidth of B_L Hz. The rapid increase in the normalized error variance as the parameter Γ is decreased represents the deterioration of the timing performance. The presence of background noise (k_n) causes the increase to occur at higher values of Γ. Since $2\beta P$ represents the rate of occurrence of signal photoelectrons during pulse transmission, the parameter Γ in (31) can be considered as an indication of the "denseness" of signal counts, indicating the accumulation of electrons over a time period equal to the reciprocal of the loop bandwidth. Alternatively, by substituting for β, we can write $\Gamma = \eta A 2P/hfB_L$, which has now the familiar interpretation as the ratio of total received pulse power to the quantum noise in the loop B_L bandwidth. The ratio $\beta k_n/B_L$ has a similar interpretation in terms of received background noise power.

It should be pointed out that the variances plotted in Fig. 6 correspond to the relatively simple tracking loop in Fig. 3. Some improvement in performance can often be attained by designing more complicated tracking systems. For example, by simultaneously processing with a parallel integrator over the leading edge of the subsequent sync pulse [i.e., over the time interval $(3W/2, 5W/2)$] one can integrate and negatively combine with $e(t)$ in Fig. 3 to strengthen (double) the error amplitude. This effectively doubles the signal power and theoretically produces a 3-dB power savings.

[3] It should be noted that [9] considered the case of phase tracking a desired sine wave with additive shot noise. The optical model here involves an input shot process with signal imbedded within, as indicated by (3). Both models lead to infinite order equations but with slightly different coefficients.

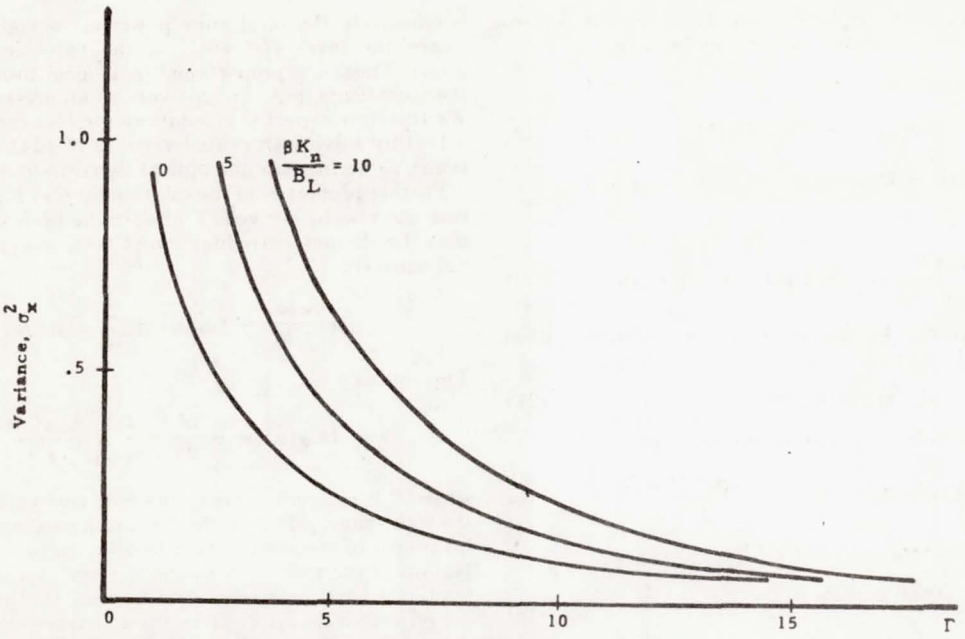


Fig. 6. Variance of normalized error τ/W versus loop signal-to-noise ratio Γ, pure sync. ($\Gamma = 2\beta P/B_L$.)

The interested reader is referred to [5] for further discussion of these possible modifications.

IV. MODULATION-DERIVED EDGE TRACKING WITH PPM

In a PPM system the received optical intensity is no longer periodic, but varies in position according to the data bit sequence. For example, in binary PPM, if the optical intensity is written as in (1), then its modulation during a bit period is given by $p(t)$ in (2) if a binary one is sent, but is given by $-p(t)$ if a binary zero is sent, as it is obvious from Fig. 2(a). A receiver attempting to attain time synchronization by edge tracking the center transitions during each bit period will be adversely affected by the data modulation. If a datum one is sent, a timing error τ will generate a loop error voltage of $R_{vp}(\tau)$, as discussed in Section III. However, if a datum zero is sent, an error voltage of $-R_{vp}(\tau)$ is generated in the same loop. Hence, for equally-likely data bits, the average error voltage within the loop is then $[R_{vp}(\tau)$ (probability of one being sent) $- R_{vp}(\tau)$ (probability of zero being sent)$] = 0$. That is, on the average, no loop error is generated for controlling the receiver timing oscillator during modulation reception.

To compensate for this modulation, an augmented edge-tracking system can be used, as shown in Fig. 7.[3] The decision loop attempts to determine the true data bit, using this decision to properly modify the sign of the loop error voltage. This can be implemented by multiplying the generated error in a delayed (by one-bit period) tracking loop by a plus or minus one, depending on the data bit. The latter decision is made from a count comparison over each possible bit subinterval as they arrive. Thus, the error in the delayed edge-tracking loop becomes $be(t)$, where

$$b = \begin{cases} +1 & \text{if one is sent,} \\ -1 & \text{if zero is sent,} \end{cases} \tag{32}$$

of equivalently,

$$b = \begin{cases} +1 & \text{if } k_1 \geq k_2 \\ -1 & \text{if } k_1 < k_2 \end{cases} \tag{33}$$

where k_1, k_2 are the counts over the first and second subinterval of each bit period. The differential equation in (8) for the tracking loop error now becomes

$$\frac{d\tau}{dt} = \frac{d\tau_1}{dt} - K[be(t)]. \tag{34}$$

[3] Again we point out that even in the modulation case more complicated tracking loops can be derived [5] which achieve performance improvement by differencing adjacent bit intervals.

Since the counts in (33) are random counts, the parameter b is a random variable. Thus, the coefficients $K_j(\tau)$ in (17) for the steady-state density will be a function of this variable, and therefore require a subsequent average over its statistics. When a one is sent the probability that $k_1 \geq k_2$ is equivalent to the probability that the one is correctly detected, whereas the probability that $k_1 < k_2$ corresponds to the probability that an error is made. Hence, when a one is sent,

$$b = \begin{cases} +1 & \text{with probability } 1 - PE_1 \\ -1 & \text{with probability } PE_1 \end{cases} \tag{35}$$

where PE_1 is the bit error probability when a one is sent. When a zero is sent, the above signs are reversed and PE_1 is replaced by PE_0. It should be remembered, however, that the timing for this subinterval counting is in turn controlled by the receiver loop timing signal, which will have loop timing errors incorporated within it. Thus, the bit-error probabilities in (38) must include these timing error effects. (A timing error between the true bit-arrival time and the start of subinterval counting will cause the counting to occur over an offset interval.) The effect of these timing errors on bit decisioning has been previously derived, [2], and a typical average bit error probability plot of $PE = \frac{1}{2}[PE_1 + PE_0]$ is shown in Fig. 8, as a function of the timing error τ, optical pulse energy $S = 2\beta PW$, and noise rate βk_n. This timing error τ is in fact a function of time, but can be considered a constant over several bit periods.

The steady-state coefficients $K_j(\tau)$ can be evaluated from (17), (19), and (34) by first conditioning on b and then averaging over the probabilities in (35). Using primes to denote the K_j coefficients when data modulation is present, and noting that $b^j = 1$ for all j even, we see that

$$\begin{aligned} K_j'(\tau) &= \begin{cases} K_j(\tau), & j \text{ even} \\ \frac{1}{2}K_j(\tau)\{(+1)[1 - PE_1] + (-1)PE_1 - (-1)[1 - PE_0] \\ \qquad - (+1)PE_0\}, & j \text{ odd} \end{cases} \\ &= \begin{cases} K_j(\tau), & j \text{ even} \\ K_j(\tau)[1 - 2PE(\tau)], & j \text{ odd} \end{cases} \end{aligned} \tag{36}$$

where $PE(\tau) = \frac{1}{2}[PE_1 + PE_0]$. Note that the dependence of PE on τ has been emphasized. The resulting steady-state density equation is again given by (16) with $K_j(\tau)$ replaced by the $K_j'(\tau)$ above. Note that the coefficients are now more complicated functions of τ due to the auxiliary decisioning, and approach the earlier results as $PE(\tau) \to 0$. In this latter case, the system is correctly identifying the true bit during each period, and essentially "removing" the binary modulation. The first coefficient,

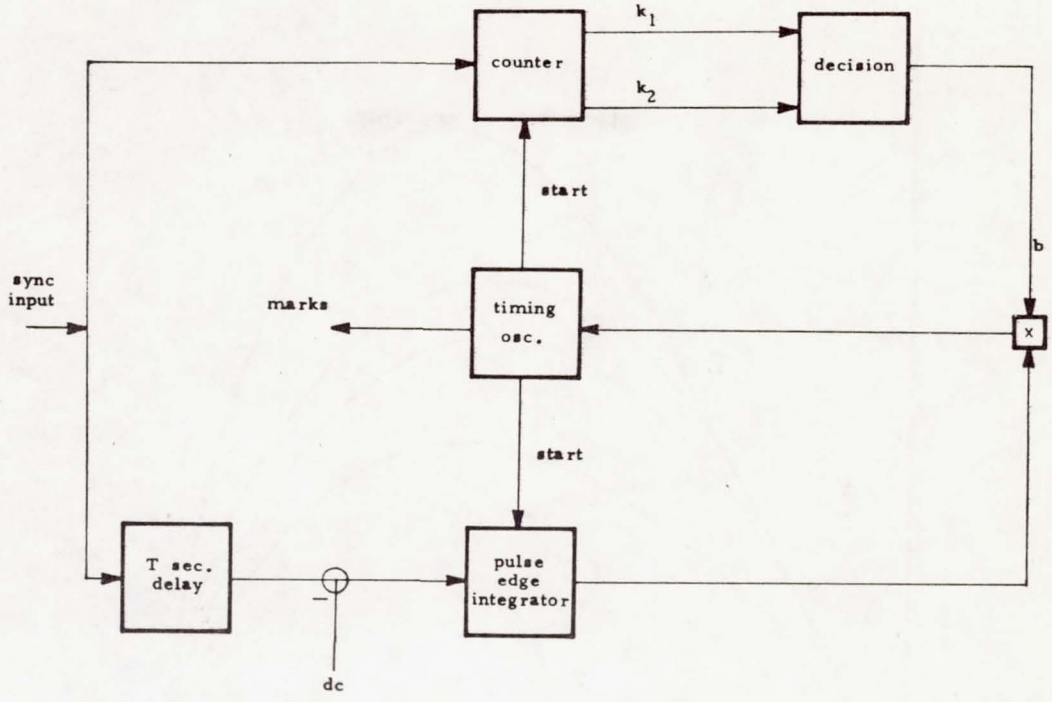


Fig. 7. Modified edge-tracking loop for PPM sync.

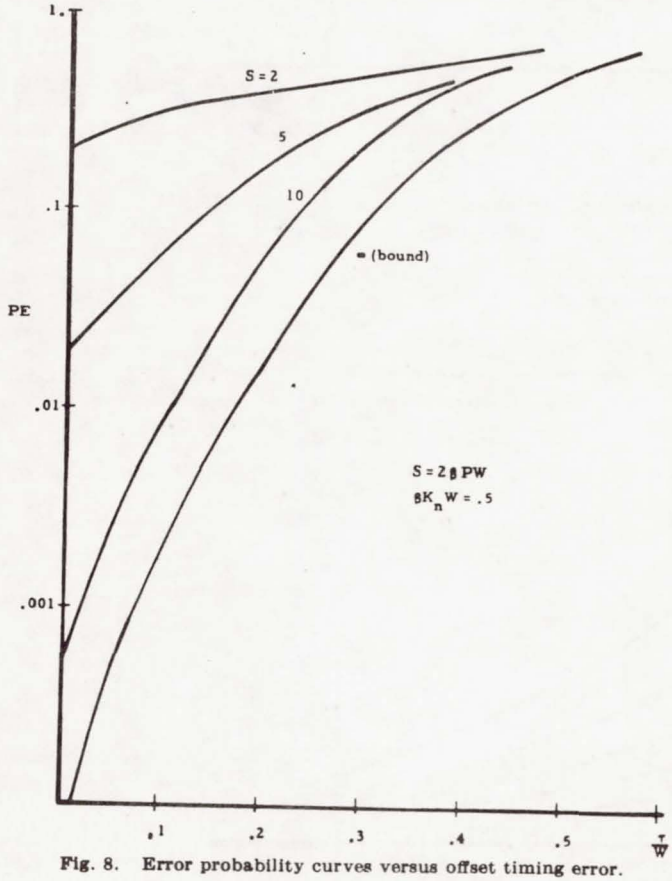


Fig. 8. Error probability curves versus offset timing error.

$$K_1'(\tau) = GEK\beta PR_{vp}(\tau)[1 - 2PE(\tau)],$$

is the average loop error function, and represents the modified nonlinearity of the mean equivalent loop, as in (12). This coefficient is plotted in Fig. 9 as a function of the normalized τ and received pulse energy S, obtained by use of Figs. 4 and 8. Note that the effect of the decision process is to reduce the width and amplitude of the tracking error function. As $PE(\tau) \to \frac{1}{2}$, there is no average error being generated for loop tracking, and the system essentially loses lock.

For the near-lock assumption [use of (23)] the previous steady-state equation is modified to

$$0 = \alpha\tau[1 - 2PE(\tau)]f(\tau) + \frac{df(\tau)}{d\tau} + \sum_{j=3}^{\infty} A_j \frac{d^{j-1}}{d\tau^{j-1}} f(\tau). \quad (37)$$

Even with this simplification, neither the solution density nor its characteristic function, can be generated as easily as in Section III, since the first coefficient is now more complicated. However, the fractional variance for this density can be estimated by approximating the coefficient $K_1'(\tau)$ in Fig. 9 by a sinusoid of proper amplitude and frequency. This latter amplitude will depend on the energy S per data bit used for decisioning, which in turn is related to the Γ parameter in (31) by

$$S = (B_L W)\Gamma \quad (38)$$

where $B_L W = B_L/2R_b = \frac{1}{2}(B_L/R_b)$. The parameter (B_L/R_b) is the ratio of tracking loop bandwidth to the data bit rate R_b, and is typically less than one. When written as in (38), $B_L W$ can also be interpreted as the fraction of the sync energy Γ appearing in the data pulse and therefore used in the auxiliary decisioning. For a fixed value of $B_L W$, each value of Γ generates a corresponding value of S, to which an effective one cycle sine wave can be fitted to the corresponding curves of $K_1'(\tau)$ as in Fig. 9. The variance can then be determined at each Γ by numerically solving a truncated version of (37), using the method discussed in Appendix B. The resulting normalized variance computed in this way is plotted in Fig. 10, as a function of Γ for several values of $B_L W$. The curve for the noiseless, pure sync operation from Fig. 6 is superimposed. The results show that a deterioration of performance occurs over pure sync, due to the decisioning process, and can therefore be considered as the price to be paid for modulation derived synchronization. Note that the decisioning causes system degradation similar to an effective loss in signal-to-noise ratio (reduced Γ) and can therefore be interpreted as a power loss in the sync subsystem.

Although the use of the curves in Fig. 10 are convenient for assessing performance, their derivation requires a somewhat lengthy calculation. Furthermore, this computation must be repeated at each desired value of background noise k_n. However a simpler method can be used, at the expense of analytical accuracy, to derive similar curves. This method makes use of a form of truncated quasi-linear solution, which basically amounts to reducing (37) to a first-order equation, and replacing the first coefficient by a modified linear coefficient as in (24), but retaining its dependence on the decision error probability PE. To accomplish this linearization we first

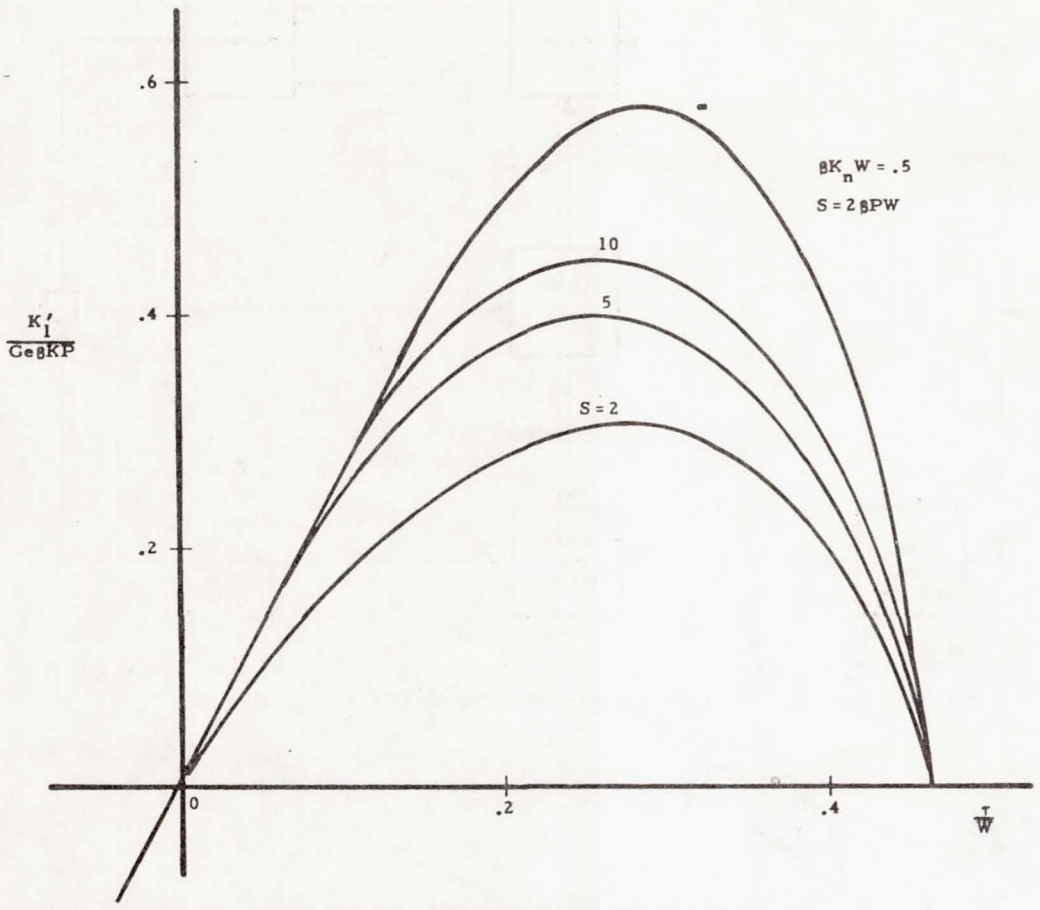


Fig. 9. Modified loop error characteristic for PPM sync.

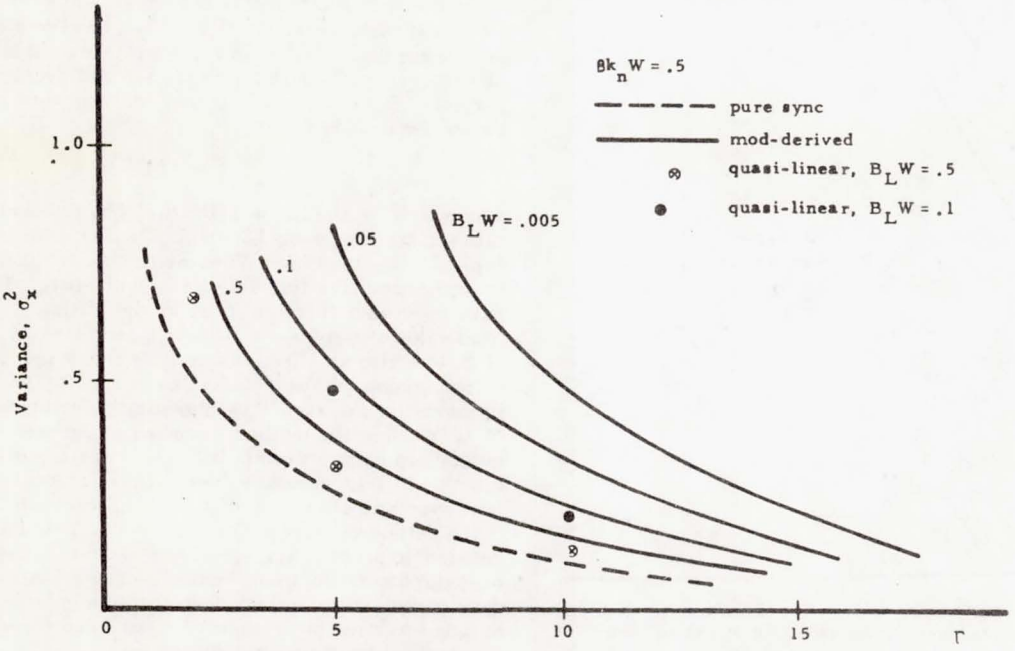


Fig. 10. Variance of normalized error versus loop signal-to-noise Γ (PPM sync).

recognize that PE depends on both pulse energy S, and timing error τ, and we write this as $PE(S,\tau)$. To linearize, we replace the functional dependence on τ by rms value of τ; i.e., $\tau_{rms} = (1/\alpha)^{1/2}$. Thus we consider instead $PE[S(1/\alpha)^{1/2}]$. The quasi-linear differential equation for the timing error density $f(\tau)$ is then taken as

$$\frac{df(\tau)}{d\tau} + \alpha\{1 - 2PE[S,(1/\alpha)^{1/2}]\}\tau f(\tau) = 0. \tag{39}$$

Note that the equation is linear in $f(\tau)$, but the coefficients are nonlinear in α. The solution for $f(\tau)$ in (39) yields a Gaussian density with normalized variance

$$\sigma_x^2 = \frac{1}{W^2\alpha\{1 - PE[S,(1/\alpha)^{1/2}]\}}. \tag{40}$$

For the shot noise limited case, $1/W^2\alpha$ is given in (30), S is related

to Γ by (38), and the variance depends only upon the parameters Γ, B_LW, and k_n/B_L. The values of PE at any value of $S = B_LW\Gamma$ and τ_{rms} can be obtained from curves similar to Fig. 8. Several points of the above variance for the shot-noise limited case are superimposed in Fig. 10. The results tend to display the same behavior for tracking performance, although the variance values indicate slightly lower variances than the more accurate results determined earlier.

APPENDIX A

The computation of the coefficients in (17) is hampered by the sampled data nature of the tracking loop in Fig. 3. We consider instead a continuous first-order loop in which the sampler is neglected, and its time averaged coefficients will be used as an approximation to the steady-state K_j. This is equivalent to assuming that the short-term integration over the pulsewidth is extremely narrow relative to the time variations of the input process, and can be neglected. The error signal $e(t)$ is therefore taken as the signal prior to the integrator, or

$$e(t) = [i(t) + i_n(t) - Ge\beta(P + k_n)]y(t - \tau_1)$$

where $y(t)$ is defined in (11). Substituting for $i(t)$ from (3), and evaluating (18) yields

$$(-\Delta\tau) = C\sum_{m=1}^{N} y(t_m - \tau_1) + K\int_t^{t+\Delta t} i_n(z)\,dz - C\beta(P + k_n)\Delta t \tag{A1}$$

where $C = GeK$ and $N = N(t,t + \Delta t)$, the latter defined in (6). To determine the K_j coefficients, the time averaged moments of $\Delta\tau$ must be calculated. The first two moments are as follows. Denoting statistical averages with overbars, we have

$$[\overline{-\Delta\tau}] = C\left[\overline{\sum_m^N y(t_m - \tau_1)}\right] + K\int_t^{t+\Delta t}[\overline{i_n(z)}]\,dz - C\beta(P + k_n)\Delta t. \tag{A2}$$

Now $i_n(z) = 0$ and it is known that Poisson shot noise has mean [11, p. 1619]

$$\overline{\sum_m^N y(t_m - \tau_1)} = \beta\int_t^{t+\Delta t} y(t_m - \tau_1)[I(t_m - \tau_1) + k_n]\,dt_m. \tag{A3}$$

In the limit as $\Delta t \to 0$

$$\lim_{\Delta t\to 0}\int_t^{t+\Delta t} y(a - \tau_1)[I(a - \tau_1) + k_n]\,da \to (\Delta t)[y(t - \tau_1)I(t - \tau_1) + k_n]. \tag{A4}$$

The above term is a function of t. The first coefficient is derived as the time-average value over an integration period W, assuming that τ does not change during this interval. Hence,

$$\begin{aligned} K_1 &= \frac{1}{W}\int_{\tau_1}^{\tau_1+W}\left\{\lim_{\Delta t\to 0}\left(\frac{\Delta\tau}{\Delta t}\right)\right\}dt \\ &= -C\frac{\beta}{W}\int_{-\infty}^{\infty} y(t - \tau_1)I(t - \tau_1)\,dt + CP\beta \\ &= -C\frac{\beta}{W}\int_{-\infty}^{\infty} y(t - \tau_1)[I(t - \tau_1) - P]\,dt. \end{aligned} \tag{A5}$$

The calculation of the mean-squared value of (A1) requires computation of the cross products involved. However, noting that the eventual computation of the K_j requires a division by Δt followed by a limit as $\Delta t \to 0$, only the terms of order Δt need be retained. In particular, we see from (A4) that any product of averages of the shot-noise summation will always be at least of order Δt^2. Hence, we have

$$\overline{(\Delta\tau)^2} = C^2\left[\overline{\sum_m^N y^2(t_m - \tau_1)}\right] + K^2\iint_t^{t+\Delta t}\overline{i_n(z_1)i_n(z_2)}\,dz_1\,dz_2 + 0(\Delta t)^2. \tag{A6}$$

The average of the first summation is known to be [11, p. 1619]

$$C^2\beta\int_t^{t+\Delta t} y^2(t_m - \tau_1)[I(t_m - \tau_1) + k_n]\,dt + 0(\Delta^2). \tag{A7}$$

Since the circuit noise is white with spectral level N_0, the second integral in (A6) is known to be $K^2N_0\Delta t/2$ [12, p. 86]. Proceeding as in (A5), we have

$$K_2 = C^2\frac{\beta}{W}\int_{-\infty}^{\infty} y^2(t - \tau_1)[I(t - \tau_1) + k_n]\,dt + K^2N_0/2. \tag{A8}$$

Similarly, we have

$$\overline{(\Delta t)^j} = (-C)^j\left[\overline{\sum_m^N y^j(t_m - \tau_1)}\right] + 0(\Delta t)^2, \qquad j \geq 3. \tag{A9}$$

By manipulating as above we finally derive

$$K_j = (-C)^j\frac{\beta}{W}\int_{-\infty}^{\infty} y^j(t - \tau_1)[I(t - \tau_1) + k_n]\,dt. \tag{A10}$$

Equations (A5), (A8), and (A10) are summarized in (19) of the paper. We point out that the first coefficient above is the same as would be calculated for the sampled system, but the higher coefficients would become extremely more difficult to determine.

APPENDIX B

The procedure here follows that of Ohlson [9]. Consider the equation

$$0 = [\alpha Q\sin 2\pi\tau]f(\tau) + \frac{df}{d\tau} + \sum_{j=2}^{N} A_j\frac{d^{j-1}f(\tau)}{d\tau^{j-1}}, \qquad |\tau| \leq W/2 \tag{B1}$$

which is a truncated version of (37) with $K_1'(\tau) = Q\sin 2\pi\tau$. We assume an even solution having the form

$$f(\tau) = \frac{1}{W}\sum_{k=0}^{\infty} C_k\cos\left(\frac{2\pi k}{W}\right)\tau, \qquad |\tau| \leq W/2 \tag{B2}$$

where

$$C_k = \int_{-W/2}^{W/2} f(\tau)\cos\left(\frac{2\pi k}{W}\right)\tau\,d\tau. \tag{B3}$$

If we substitute (B2) into (B1), collect harmonic terms, set the resulting coefficients equal to zero, we derive the following second-order recursive equations among the C_k:

$$C_{k+1} = C_{k-1} + \left[-\frac{2k}{Q\alpha} + \sum_{\substack{j=2\\(\text{odd})}}^{N}\frac{2A_jk^j}{Q\alpha}\right]C_k. \tag{B4}$$

The above allows a generation of all subsequent C_k from the first two, C_0 and C_1. These latter two are found from the conditions that 1) $f(\tau)$ be a probability density over $(-(W/2),W/2)$ and 2) for large α, $f(\tau)$ must approach the known solution corresponding to $A_j = 0$ in (40). From (B3) we see that the first condition requires that $C_0 = 1$, while for $A_j = 0$, (B4) becomes

$$C_{k+1} = C_{k-1} - \left(\frac{2k}{Q\alpha}\right)C_k, \qquad C_0 = 1. \tag{B5}$$

The solution is then

$$C_k = \frac{I_k(Q\alpha)}{I_0(Q\alpha)} \tag{B6}$$

where I_k is the imaginary Bessel function of order k. Thus, C_1 was selected as $I_1(Q\alpha)/I_0(Q\alpha)$ in subsequent analysis using (B4). The density $f(\tau)$ can now theoretically be constructed by solving (B4) and using the C_k in (B2).

We are primarily interested in the normalized variance of this tracking error, given by

$$\begin{aligned} \sigma^2 &= \frac{1}{W^2}\int_{-W/2}^{W/2}\tau^2 f(\tau)\,d\tau \\ &= \frac{C_0}{12} + \sum_{k=1}^{\infty} C_k\int_{-1/2}^{1/2} x^2\cos(2\pi kx)\,dx \\ &= \frac{C_0}{12} + \sum_{k=1}^{\infty} C_k\left[\frac{(-1)^k}{2\pi^2k^2}\right]. \end{aligned} \tag{B7}$$

The above can be computed directly from the C_k generated from (B4). To examine the truncation error (B4) was numerically solved for Q in the range (0.1–0.5) and $N = 3$ and 5. For the range of interest ($1 \leq \alpha \leq 100$) no noticeable change in variance appeared for N greater than 3. Hence, the truncation was limited to $N = 3$ in all subsequent results. With $N = 3$, the variance was then computed from (B7) as a function of α, after generation of the C_k from (B4). The value of Q, which itself depends on α, was determined for each α from the curves of Fig. 9. The resulting variance is plotted in Fig. 10 of the paper, assuming shot-noise limited operation.

REFERENCES

[1] S. Karp and R. M. Gagliardi, "The design of a pulse-position modulated optical communication system," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 670–676, Dec. 1969.

[2] R. M. Gagliardi, "The effect of timing errors in optical digital systems," *IEEE Trans. Commun.*, vol. COM-20, pp. 87–93, Apr. 1972.

[3] J. Stiffler, *Theory of Synchronous Communication*. Englewood Cliffs, N. J.: Prentice-Hall, 1971, pt. 2.

[4] W. C. Lindsey, *Synchronization Systems in Communications and Control*. Englewood Cliffs, N. J.: Prentice-Hall, 1971.

[5] W. Lindsey and M. Simon, *Telecommunication Systems Engineering*. Englewood Cliffs, N. J.: Prentice-Hall, 1973, ch. 9.

[6] R. M. Gagliardi and S. Karp, "M-ary Poisson detection and optical communications," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 208–216, Apr. 1969.

[7] J. R. Clark, "Estimation for Poisson processes with applications in optical communications," Ph.D. dissertation Mass. Inst. Tech., Cambridge, Mass., Sept. 1971.

[8] D. L. Snyder and I. B. Rhodes, "Phase and frequency tracking accuracy in direct-detection optical communication systems," *IEEE Trans. Commun.*, vol. COM-20, pp. 1139–1142, Dec. 1972.

[9] J. E. Ohlson, "Phase-locked loop operation in the presence of impulsive and Gaussian noise," *IEEE Trans. Commun.*, vol. COM-21 pp. 991–996, Sept. 1973.

[10] W. C. Lindsey and M. K. Simon, "Data-aided carrier tracking loops," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 157–168, Apr. 1971.

[11] S. Karp, E. L. O'Neill, and R. M. Gagliardi, "Communication theory for the free-space optical channel," *Proc. IEEE*, vol. 58, pp. 1611–1626, Oct. 1970.

[12] A. Viterbi, *Principles of Coherent Communications*. New York: McGraw-Hill, 1966.